The Workshops of the Eleventh International AAAI Conference on Web and Social Media
AAAI Technical Report WS-17-17: News and Public Opinion

The Impact of News Values and Linguistic Style on the Popularity of Headlines on Twitter and Facebook

Alicja Piotrkowicz
Vania Dimitrova
School of Computing
University of Leeds, UK
{scap,V.G.Dimitrova}@leeds.ac.uk

Jahna Otterbacher
Social Information Systems
Open University of Cyprus
jahna.otterbacher@ouc.ac.cy

Katja Markert
Institute of Computational Linguistics
Heidelberg University, Germany
markert@cl.uni-heidelberg.de

Abstract

A large proportion of audiences read news online, often accessing news articles through social media like Facebook or Twitter. A distinguishing characteristic of news on social media is that the most prominent (and often the only visible) part of the news article is the headline. We investigate the impact of headline characteristics, including journalistic concepts of *news values* and linguistic *style*, on the article's social media popularity. Using a large corpus of headlines from *The Guardian* and *New York Times* we derive these features automatically and correlate with social media popularity on Twitter and Facebook. We found most of them to have a significant effect and that their impact differs between the two social media and between news outlets. Further investigation with a crowdsourced study confirms that news values and style influence the audiences' decisions to click on a headline.

Introduction

A vast amount of digital content, ranging from news (92,000 articles published online daily[1]), blogs (nearly 70 million Wordpress posts published each month[2]), forum posts (over 73 million Reddit submissions in 2015[3]), and videos (300 hours of video uploaded every minute on YouTube[4]) features a headline (i.e. a title). Titles play a vital role in attracting audiences' attention to online artefacts, which is crucial when the artefacts are accessed via social media.

In this paper, we focus on news article headlines to investigate whether headline characteristics have impact on the popularity of news items shared via social media. We have selected popular broadsheet news outlets, whose presence in the social arena is vital given the growing importance of social media as a source of news and the increasing appearance of fake news in the social space. Social media sites, such as Facebook and Twitter, are becoming the main medium for news consumption. According to the latest report by Pew Research Center, 62% of the Americans get news on social media[5]. BBC reports that social media is young people's main source of news[6], which makes social media crucial to reach that demographic.

In this paper we investigate the impact of two types of automatically extracted headline features on social media popularity of news articles. Firstly, we draw insights from journalism literature, which postulates the existence of *news values* — aspects of an event that determine whether and to what extent it is reported in news outlets. Secondly, we investigate the impact of linguistic style (i.e. phrasing) on news article popularity in considerably more depth than previously done in computational models of news articles.

The approach presented in this paper can be used by computational journalism researchers to derive insights about news audiences and news consumption. The investigation of style features can also be very informative for headline generation research.

Related Work

Headlines hold a special place in news discourse. Researchers have looked at their functions (Althaus, Edy, and Phalen 2001) and the role their play for information processing (Dor 2003). Considerable attention has been paid to the language of headlines, in particular what sets them apart from other text types, e.g. passive transformation and nominalisation (Fowler 1991, p.77-79), deliberate ambiguity (Brône and Coulson 2010), and untensed verbs (Chovanec 2014). Some researchers worked on identifying how higher level concepts, such as sensationalism (Molek-Kozakowska 2013), or click-bait (Blom and Hansen 2015), are captured in the headline language. In our work we analyse the headline language directly (style features) and the linguistic expression of higher level concepts (news values features).

For the news values features we take inspiration from the news selection literature. The seminal work in this area is by Galtung and Ruge (1965) who propose a set of so-called news values as event properties which determine the space afforded to an event by news outlets. Other empirical work in the journalism community, using manual content analysis, confirmed the importance of news values for news selection (Kepplinger and Ehmig 2006). In contrast, our aim is to explore the impact of news values on social media *reception*. A variety of news values taxonomies have been

[1]http://bit.ly/2cPaQZK
[2]https://wordpress.com/activity/
[3]https://redditblog.com/2015/12/31/reddit-in-2015/
[4]http://www.statisticbrain.com/youtube-statistics/
[5]http://pewrsr.ch/27TOfhz
[6]http://www.bbc.com/news/uk-36528256

proposed, including: Gans (1979), Bell (1991), Harcup and O'Neill (2001), and Bednarek and Caple (2012). There is, however, considerable agreement between these taxonomies (Caple and Bednarek 2013), allowing us to concentrate on the news values which are most frequently mentioned.

The second feature group – linguistic style – refers to wording and is largely topic-independent. Linguistic style has been previously used to model the popularity of short texts, such as Reddit image titles (Lakkaraju, McAuley, and Leskovec 2013) and tweets (Tan, Lee, and Pang 2014). Some commonly used features include: ratios for various parts of speech, sentiment, and similarity to a language model. Headline text is a similar phenomenon, which brings its unique challenges. For example, since headlines offer limited context, word-level sentiment rather than context-sensitive state-of-the-art methods is often used (Tan, Lee, and Pang 2014; Gatti et al. 2016; Szymanski, Orellana-Rodriguez, and Keane 2016). We adapt these frequently used features to work with headlines. We also draw from the news domain to implement features which were not explored for this task before. This includes syntactic and lexical simplicity (mentioned in news selection literature, e.g. Bell (1995)) and implemented using features drawn from NLP research on readability (Pitler and Nenkova 2008; Feng et al. 2010; Kate et al. 2010). Following advice on writing headlines from news style guides (e.g. *The Guardian*'s style guide[7]) we use punctuation and ambiguity.

Data Collection

Our datasets consist of headlines from two news outlets and their social media popularity.

News Corpora

We collected headlines from two major broadsheet newspapers — *The Guardian* (Guardian) and *New York Times* (NYT). We downloaded all headlines with their associated metadata published during April 2014 (Guardian training; 11,980 articles), July 2014 (Guardian test; 13,806 articles), October 2014 (NYT training; 5074 articles), and December 2014 (NYT test; 5011 articles)[8]. Table 1 includes some example headlines and their social media popularity.

Preprocessing. All headlines were part-of-speech tagged (Stanford POS Tagger (Toutanova et al. 2003)) and parsed (Stanford Parser (Klein and Manning 2003)). We used wikification (a method of linking keywords in text to relevant Wikipedia pages) to identify a broad range of entities. Headlines were wikified using the TagMe API[9], a tool meant for short texts, making it suitable for headlines.

Social Media Popularity

We define the popularity of a news article as the number of times it is cited on Facebook and Twitter. We adopt metrics used in earlier research. Number of tweets (Twitter) has been previously used to predict the popularity of news articles (Bandari, Asur, and Huberman 2012), and political parties (Tumasjan et al. 2010); and number of likes (Facebook) has been used to measure popularity of brand pages (De Vries, Gensler, and Leeflang 2012).

Collection and validation. The article URL was used as the search query for the Twitter Search API[10] to obtain the number of tweets and retweets. We queried the Twitter API one, three, and seven days after the article's publication. The process was repeated for Facebook likes and shares using the Facebook FQL API.[11] Because the APIs return a sample of all results, we checked the correlations for 100 articles between the citations we collected via the API and the number of citations that appear on *The Guardian* article website[12]. The correlation is over 0.95 for all measures.

Popularity measures. Tweets and retweets, as well as shares and likes, are combined into two metrics: Twitter and Facebook popularity. It has been previously suggested that it takes approximately four days for a message to propagate through social media (Leskovec, Backstrom, and Kleinberg 2009). We found that in our datasets Twitter and Facebook popularity after three and seven days did not differ significantly, and so throughout the paper we report popularity after three days, yielding two social media popularity measures: T = Twitter popularity after three days, and F = Facebook popularity after three days. Considering both direct (tweets, shares) and indirect (retweets, likes) citations allows to investigate the *overall social media popularity*.

Headline Features: Implementation and Correlations with Popularity

We developed a novel approach to analysing headlines using automatically extracted news values and style features. We present (i) summaries of feature implementations, and (ii) results of an exploratory study that investigates how the extracted features correlate with news articles' Twitter and Facebook popularity (Table 2). For numeric features we used Kendall's τ correlation. For binary features using the Wilcoxon signed rank test we compared whether the feature median is significantly higher or lower than the overall median for that social media popularity measure.

News Values

We use six news values which have been frequently included in taxonomies (cf. Related Work). These have been previously implemented and validated in Piotrkowicz, Dimitrova, and Markert (2017). When looking at the impact of news values on popularity, we consider two conditions: 'hard news' (the genre for which news values have originally been theorised) and all genres (to test whether news values work across all article genres).

Prominence: Mention entities that are popular. Reference to prominent entities is one of the key news values. Prominence can be viewed as importance or recognisability.

[7]http://www.theguardian.com/guardian-observer-style-guide-a

[8]*The Guardian* data: Guardian Content API (http://www.theguardian.com/open-platform), *New York Times*: NYT Article Search API (http://developer.nytimes.com/docs).

[9]http://tagme.di.unipi.it/

[10]https://dev.twitter.com/docs/api/1.1/get/search/tweets

[11]https://developers.facebook.com/docs/technical-guides/fql/

[12]*New York Times* does not provide this data for their articles.

Table 1: Examples of most and least popular headlines.

	The Guardian	*New York Times*
Most popular	”Nobel winner Gabriel García Márquez hospitalised in Mexico City” (T=299, F=38566) “Facebook app revealed to be cause of iPhone battery woes” (T=923, F=5657)	“From a Rwandan Dump to the Halls of Harvard” (T=3442, F=69285) ”Malala Yousafzai, Youngest Nobel Peace Prize Winner, Adds to Her Achievements and Expectations” (T=245, F=21755)
Least popular	“What’s next after the iGeneration?” (T=0, F=0) “Tunnel vision on rail competition” (T=7, F=0)	“The Risina Bean Is Worth the Hunt” (T=0, F=1) “Home Sales Around the Region” (T=0, F=17)

Implementation. We approximate Prominence as the amount of online attention an entity gets and implement six features. We extend previous work by using wikification for obtaining entities, which ensures a wide variety of entity types. The first feature is the number of entities identified in a headline. To gauge the online attention that an entity gets we consider two sources: entity mentions in headlines and Wikipedia page views. The news recent prominence feature is the number of entity mentions in headline of the relevant news outlet (*The Guardian* or *New York Times*). Since we are using wikification an entity corresponds to a Wikipedia page title. Using Wikipedia pageviews we look at long-term prominence (the median number of daily page views over a year for an entity) and day-before prominence (number of page views for an entity the day before a given headline was published). Finally, we consider how some entities have a bursty, i.e. non-steady, prominence. An entity is defined as *being in a burst* if its moving average of Wikipedia page views in a given time frame is above a predefined cut-off point (the burst detection algorithm we use has been adapted from Vlachos et al. (2004)). We implement two features – the size of the current burst (how much above the cutoff point are an entity’s page views) and burstiness (how many times over a year an entity has been in a burst). Since a given headline may feature multiple entities, all prominence measures are aggregated via summation over all entities.

Impact on popularity. Most Prominence features are positively correlated with popularity on Twitter and Facebook. Number of entities is positively correlated for Guardian, but negatively for NYT. Entities prominence in headlines is more highly correlated for Twitter, but in case of Facebook it is higher for NYT than for Guardian. Wikipedia long-term and day-before prominence are among the best correlated features – especially for Twitter and when looking at all genres. Current burst size is only significantly correlated for NYT, whereas burstiness is more highly correlated for Twitter measures. Overall, Prominence features have higher correlations with Twitter measures than with Facebook; and for features using Wikipedia they are higher for all genres setting, rather than the news subset.

Sentiment: More negative and biased vocabulary helps. Sentiment-charged language has been highlighted for news discourse (Bednarek and Caple 2012). Features relating to sentiment and emotionality have been shown to influence a news article’s virality (Berger and Milkman 2012). The effect of sentiment on headline popularity has been studied by Reis et al. (2015).

Implementation. Since headlines offer limited context, sentiment analysis carried out on word-level is frequently used (cf. Tan, Lee, and Pang (2014), Gatti et al. (2016), Szymanski, Orellana-Rodriguez, and Keane (2016)). We use SentiWordNet (Baccianella, Esuli, and Sebastiani 2010) positivity and negativity scores of a headline’s content words to combine them into sentiment and polarity scores (Kucuktunc et al. 2012). We also look at indirect sentiment expressed via connotations and biased words. For example, a word may be in itself objective, but carry a negative connotation (e.g. *scream*). Using a connotations lexicon (Feng et al. 2013) we calculate the proportion of connotated words (positive or negative) to all content words. Secondly, using a bias lexicon (Recasens, Danescu-Niculescu-Mizil, and Jurafsky 2013) we calculate the percentage of biased content words. For example, the same political organisation can be described as *far-right*, *nationalist*, or *fascist*, each of these words indicating a bias towards a certain reading.

Impact on popularity. The sentiment is only significant for all genres in Guardian dataset. The negative correlation means that headlines with higher negativity are more popular which follows the news values literature (Galtung and Ruge 1965). Polarity (here, sentiment-charged language overall) is positively correlated across all measures, and the highest for NYT all genres. The connotations feature is significantly correlated for most measures (excluding NYT news subset), while bias is significantly correlated in all cases, but is slightly higher for NYT. Overall, sentiment-charged and biased language in headlines has a positive impact on social media popularity, which is in line with previous research (Louis and Nenkova 2013; Lakkaraju, McAuley, and Leskovec 2013; Tan, Lee, and Pang 2014).

Magnitude: Enhance or diminish the events you’re describing. The size, magnitude, or impact of the event (Johnson-Cartee 2005; Harcup and O’Neill 2001) is considered to influence news selection.

Implementation. To ensure that the feature implementation remains topic-independent, we focus on explicit linguistic indicators of event size: comparatives and superlatives (indicated by part-of-speech tags), and amplifiers (intensifiers and downtoners). We obtain wordlists of 248 intensifiers and 39 downtoners by combining the lists given in Quirk et al. (1985) and Biber (1991). Counts of these features are averaged using the number of words in a headline.

Impact on popularity. The Magnitude features are significantly correlated with most measures (excepting NYT news subset). The highest correlations in this feature group are

Table 2: Feature correlations (Guardian = *The Guardian*, NYT = *New York Times*, News = news subset, All = all genres). Numeric features: Kendall's τ (* $p<0.05$, ^ $p<0.01$); binary features: feature median is higher/lower (↑ / ↓ $p<0.05$) than the overall median (in brackets in column headings) using Wilcoxon signed rank test.

		Guardian-News		Guardian-All		NYT-News		NYT-All	
	Feature name	T (57)	F (83)	T (41)	F (42)	T (117)	F (200)	T (102)	F (153)
Prominence	number of entities	0.04^	0.01	0.07^	0.03^	-0.06^	-0.06^	-0.02	-0.01
	News recent prominence	0.05^	0.01	0.11^	0.02^	0.11^	0.07^	0.11^	0.06^
	Wikipedia long-term prominence	0.08^	0.05^	0.16^	0.11^	0.1^	0.05^	0.11^	0.08^
	Wikipedia day-ago prominence	0.08^	0.05^	0.15^	0.11^	0.12^	0.07^	0.13^	0.09^
	Wikipedia current burst size	0	-0.01	0	0.01	0.06^	0.06^	0.08^	0.08^
	Wikipedia burstiness	0.05^	0.02	0.07^	0.02^	0.04^	0.01	0.06^	0.03^
Sentiment	sentiment	-0.02	0.01	-0.06^	-0.04^	0	0.02	0.01	0.02
	polarity	0.09^	0.06^	0.1^	0.09^	0.09^	0.1^	0.11^	0.12^
	connotations	0.06^	0.04^	0.05^	0.06^	0.01	0.02	0.05^	0.06^
	bias	0.05^	0.03*	0.07^	0.06^	0.08^	0.07^	0.09^	0.08^
Magnitude	comparative/superlative	0.06^	0.03*	0.03^	0.03^	0.02	0.02	0.04^	0.03^
	intensifiers	0.06^	0.05^	0.04^	0.03^	0.01	0.02	0.03^	0.04^
	downtoners	0.05^	0.07^	0.03^	0.02^	0.03*	0.01	0.04^	0.03*
Proximity	proximity	35↓	38↓	40	34↓	39↓	38.5↓	40↓	40↓
Surprise	surprise	-0.04^	-0.03*	-0.02^	-0.01	-0.02	-0.02	-0.01	-0.01
Uniqueness	headline uniqueness	-0.06^	-0.04^	-0.06^	-0.08^	0.02	-0.01	0.01	-0.02
Brevity	number of words	0.13^	0.12^	0.14^	0.11^	0.16^	0.1^	0.2^	0.15^
	number of characters	0.09^	0.07^	0.13^	0.09^	0.15^	0.08^	0.19^	0.12^
Simplicity	parse tree height	0.09^	0.08^	0.15^	0.12^	0.13^	0.11^	0.16^	0.13^
	number of tree nodes	0.11^	0.1^	0.15^	0.12^	0.15^	0.11^	0.17^	0.15^
	entropy	-0.08^	-0.1^	-0.07^	-0.1^	0	-0.07^	0.03^	-0.04^
	proportion of difficult words	-0.05^	-0.04^	-0.06^	-0.06^	-0.07^	-0.03^	-0.06^	-0.02
	information content	0.03*	0.02*	0.09^	0.07^	0.05^	0.04^	0.02*	0
	word frequency	-0.05^	-0.03^	0	-0.03^	0	-0.02*	0.01	-0.02*
Unambiguity	number of senses	0.03*	0.03*	0.01*	0	0.01	0	-0.01	-0.01
	modality	43↓	39↓	57↑	64↑	30↓	25↓	46↓	42↓
Punctuation	exclamation mark	35	21↓	20↓	33	26↓	9↓	31↓	63.5
	question mark	41↓	52↓	50↑	69↑	39↓	35↓	37↓	58↓
	quote marks	37↓	43↓	49↑	68↑	41↓	32↓	41↓	35↓
Nouns	three consecutive nouns	38↓	42↓	40	35↓	38↓	38↓	40↓	42↓
	NP count	-0.11^	-0.09^	-0.14^	-0.1^	-0.11^	-0.13^	-0.12^	-0.14^
	proportion of nouns	-0.02	-0.06^	0.03^	0.03^	0.05^	0.02	0.04^	0.03^
	proportion of proper nouns	0.07^	0.06^	0.09^	0.1^	0.06^	0.03^	0.05^	0.04^
Verbs	VP count	0.11^	0.08^	0.13^	0.09^	0.07^	0.1^	0.1^	0.12^
	proportion of verbs	0.08^	0.06^	0.13^	0.1^	0.08^	0.09^	0.11^	0.11^
Adverbs	proportion of adverbs	0.09^	0.11^	0.04^	0.04^	0.06^	0.06^	0.06^	0.07^

for the Guardian news subset. It is interesting to note that relatively infrequent lexical items such as intensifiers and downtoners have an impact on the popularity of headlines. In particular, it is somewhat surprising that the presence of downtoners, which function to diminish the words they describe, also correlates positively with popularity.

Proximity: Mentioning entities geographically close to the news source is not crucial. Both geographical (Johnson-Cartee 2005, p.128) and cultural proximity (Galtung and Ruge 1965; Gans 1979) has been considered. Proximity has been linked to relevance; the justification being that news events which are geographically or culturally close to the reader are more newsworthy. We assume that readers from the same country as the news outlet constitute a large part of its readership, and so we look only at geographic proximity to the news source.

Implementation. A binary feature indicates whether a headline mentions an entity that is geographically close to the news source. We manually create two wordlists by combining names for the country, regions/states, capital city and arrive at 17 UK-related terms for *The Guardian*, and 61 US-related terms for *New York Times*). We match the keywords from the wordlists against both headline text (e.g. "*London* smog warning as Saharan sand sweeps southern *England*") and the names of Wikipedia categories of each entity supplied in the TagMe output (e.g. category POSTAL SYSTEM OF THE UNITED KINGDOM for headline "Undervaluing *Royal Mail* shares cost taxpayers 750m in one day").

Impact on popularity. Headlines which include a reference to location relevant to a given news outlet's country of origin have a significantly lower median popularity for nearly all measures. This might be due to readers being more oriented towards global news in general, or to our implementation of proximity in relation to the news source rather than the user. This latter will be addressed in our future work.

Surprise: A surprising phrasing sometimes helps. Events which involve "surprise and/or contrast" (Harcup and O'Neill 2001) make news.

Implementation. We focus on surprising phrasing in headlines (e.g. "Beekeeper creates *coat of living bees*"). We measure it by calculating the commonness of phrases in a headline with reference to a large Wikipedia corpus[13]. We first extract syntactic chunks of following types: SUBJ-V, V-OBJ, ADV-V, ADJ-N, N-N; and then obtain a count from the corpus for the phrase as well as its inflected forms (e.g. man drinks: man drinks, man drank, etc.). We then sum the counts for each phrase and calculate its log-likelihood (LL). Since we are looking for the most surprising phrase, the feature value is given as the lowest LL in the headline.

Impact on popularity. Surprising word combinations (i.e. those with lower log-likelihood) have a significant correlation for Guardian datasets, but not for *New York Times*. (Louis and Nenkova 2013), who used a similar approach to find 'creative language', also found that uncommon word combinations positively influence popularity.

Uniqueness: Similarity to headlines from recent past usually has negative impact. News has to be new – "any new comment or circumstance [...] adds to the debate" (Conley and Lamble 2006). An analysis of several storylines in our dataset showed that of two very similar headlines, the latter tends to be less popular.

Implementation. For a given headline we select past headlines from 72 hours before its publication and which have at least one TagMe entity overlapping or neither has any entities. Entity overlap helps with finding headlines that are part of the same storyline, whereas including headlines with no entities ensures better coverage. For a pair of headline and past headline vectors (created using a *tf-idf* weighted Gigaword corpus) we calculate cosine similarity. The highest cosine similarity is assigned as the feature value.

Impact on popularity. This feature significantly correlates with all Guardian measures, but not with NYT. The significant correlations are negative, which means that if there's a very similar headline in recent past, then the current headline will be less popular. This was also found to be the case for titles on Reddit (Lakkaraju, McAuley, and Leskovec 2013).

Style

We identified several stylistic aspects often highlighted in news selection literature and also investigated linguistic recommendations from news outlet style guides. We found that many aspects of phrasing investigated in the NLP community, which we include here, fall under one of the terms used in the journalism literature.

Brevity: Longer text is better. Traditionally for newspapers, space is of the essence (Bell 1995). This has also been suggested for headlines (Dor 2003). We implement Brevity as the number of tokens and the number of characters.

Impact on popularity. Both features are positively correlated with popularity, which goes against the suggestions in literature. The correlation is particularly strong for *New York Times*, especially Twitter. This might be partly due to genre effect, where genres like hard news and opinion have longer headlines that describe events, whereas regular features (crosswords, corrections, TV listings) tend to be brief.

Simplicity: Use simple vocabulary, but more complex syntax. One of the objectives of news writing is "ease of comprehension" (Cotter 2010). This can be related to vocabulary and syntax (Bednarek and Caple 2012).

Implementation. We measure syntactic complexity with parse tree height and number of non-terminal tree nodes. To measure lexical complexity we implement four features. The first two - entropy and proportion of difficult words - are obtained from a trigram language model.[14] We define a difficult word as any word not occurring among the 5000 most common words in the language model. The third lexical complexity feature is median word frequency, obtained using the unlemmatised word frequency lists.[15] The final feature is median information content calculated for nouns and verbs on British National Corpus.

[13]http://www.nlp.cs.nyu.edu/wikipedia-data

[14]The language model was built using the CMU-Cambridge Toolkit (Clarkson and Rosenfeld 1997) on the *New York Times* section of the Gigaword corpus (Graff et al. 2003).

[15]http://www.wordfrequency.info/; British National Corpus for Guardian, Corpus of Contemporary American English for NYT.

Impact on popularity. Most of the simplicity features are significantly correlated. The syntactic complexity features have relatively high correlations for Twitter in both news sources, and NYT correlations are slightly higher than those for Guardian. Surprisingly, syntactically more complex headlines seem to be preferred. Although some NLP research on readability (Pitler and Nenkova 2008) found that text length and parse tree height are negatively correlated with readability, in case of headlines it makes sense that these features (along with proportions of verbs and VPs) would be positively correlated, as longer headlines with full sentences instead of 'headlinese' would be easier to understand. Lexical features have slightly higher correlations for Guardian, the negative correlation follows literature (headlines with simpler vocabulary tend to be more popular).

Unambiguity: Avoid multiple meanings and modal events. News writing is expected to be unambiguous (Galtung and Ruge 1965). We focus on how ambiguity can be realised through lexical and syntactic means.

Implementation. We calculated the median number of senses per word from WordNet – more senses per word indicate a higher chance for ambiguity. We used the TARSQI toolkit[16] to indicate if there is a modal event (*should*) or a modal relation between events ("Gove *promises* to *abolish* illiteracy [...]") in the headline.

Impact on popularity. Number of senses is not significant for most measures, but is weakly positively correlated for three Guardian measures. Headlines with modal events tend to have a lower median popularity for most measures, however in case of all genres in the Guardian modality has a higher median popularity.

Punctuation: Avoid ! ? "" in most cases. *The Guardian*'s style guide for headlines[17] cautions against using quote, question, and exclamation marks. We implement binary features which check for their presence.

Impact on popularity. In most cases the presence of these punctuation marks results in a significantly lower median popularity. However, there are some cases where it seems to have a positive effect – namely, Guardian all genres for question and quote marks. Presence of quote marks can indicate bias (e.g. "Spanish celebrate 'conquest' of French politics"), which is consistent with the positive correlation of the bias feature for that data subset. Question marks can indicate event uncertainty (e.g. "Is the 'cost of living crisis' over?"), which follows the results for the modality feature (where the feature median was also significantly higher).

Nouns: Avoid consecutive nouns, but proper nouns are fine. *The Guardian*'s style guide discourages using too many successive nouns (so-called 'headlinese', e.g. "New York assault weapons ban"). As aspects of 'headlinese', we consider the number of consecutive nouns and the overall proportion of nouns to all words. The Yahoo guidelines for headlines (Barr 2010) recommend usage of proper names. We implement this as the proportion of proper names to all nouns.

Impact on popularity. Both the presence of three or more consecutive nouns (indicating 'headlinese') and the NP count show quite strong negative correlations (slightly lower for NYT). Surprisingly, the proportions of nouns and proper nouns seem to be positively correlated in most cases. Proper nouns indicate entities and since prominence seems to play an important role, then the positive correlation for the proportion of proper nouns could be attributed to that.

Verbs: Using verbs helps. The usage of verbs is encouraged in headlines in *The Guardian*'s style guide. We implement two features: the number of verb phrases and the proportion of verbs to all words.

Impact on popularity. Both features are positively correlated for all measures. In the Guardian dataset, correlations with Twitter are slightly higher.

Adverbs: Using adverbs helps. Adverbs, especially adverbs of manner, appear often in headlines (Bednarek and Caple 2012). We use the proportion of adverbs as a feature.

Impact on popularity. Higher proportion of adverbs has a positive impact on popularity, in particular for the Guardian news subset.

Discussion

Overall, we found that each of the extracted features is significantly correlated with at least one of the measures. This shows that headline content influences the popularity of news articles on social media.

Differences between Twitter and Facebook. The correlations with social media popularity for Twitter are slightly stronger than for Facebook. When taking into account news sources as well, in *The Guardian* dataset correlations are higher for Twitter compared to Facebook, however in *New York Times* Facebook usually has higher correlations. This variation aligns with reports which describe differences in demographics of news readers on these two websites[18]. This calls for further work which considers user demographics.

Differences between news sources. When comparing feature correlations between *The Guardian* and *New York Times*, we found that there are some features which play a particularly significant role. For example, most news values and verb-related features were more strongly correlated for Guardian; whereas for NYT it was some Sentiment features, Brevity and syntactic Simplicity. Running correlations on headlines corpora from other broadsheets would determine which news values and style features generalise across the broadsheet genre and which have a particularly significant effect only for certain news outlets. Similarly, corpora consisting of headlines from other types of news outlets (e.g. tabloids) would enable a cross-genre comparison.

Scope. Our approach is limited to features extracted directly from headline text. We do not take into consideration any external factors like visual presentation (e.g. whether the headline was displayed on the top of the page), or social effects (e.g. whether the headline tweeted by a high-profile celebrity). These confounding factors might explain the relatively low correlations. However, we believe that our features would enrich any approach which models the popularity of headlines. Future work will address such hybrid

[16]http://www.timeml.org/site/tarsqi/toolkit/index.html

[17]http://www.theguardian.com/guardian-observer-style-guide-h

[18]http://pewrsr.ch/27TOfhz

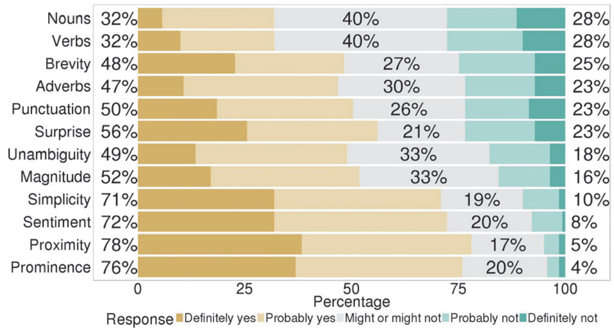

Figure 1: Survey results to the question *"I personally consider this feature when clicking on headlines"* (N=141). Percentages show the aggregated results for positive, neutral, and negative responses.

approaches, which combine content (headline text) and context (presentation, social graph) features.

Human Judgement about Headline Features

In order for a headline to gain popularity on social media, it first needs to attract the readers' attention. To investigate whether news readers themselves think these factors influence them when choosing headlines, we conducted a survey using CrowdFlower[19]. The survey presented 12 headline features (cf. Table 2; we excluded Uniqueness since that would require presenting participants with past headlines for comparison), each with a short definition and examples drawn from out datasets. For each headline feature participants were asked *"I personally consider this feature when clicking on headlines"* and given five Likert scale responses. (cf. Figure 1).

We collected nearly 200 responses. As quality control, we removed any responses where more than 75% of answers were neutral, as well as responses where time to complete was in the bottom quartile (to ensure that participants had taken time to understand the concepts). After the quality control measures, 141 responses were selected: 92 participants were 34 or younger and 49 were 35 or older; 30 were female, 111 were male; 44 were native English speakers and 97 were non-native English speakers; 96 participants read news daily, 45 weekly.

Results of the survey are presented in Figure 1. Most features had a positive response, with news values of Proximity (78%), Prominence (76%), and Sentiment (72%) receiving the highest proportions of positive responses. On the other hand, some style features like Nouns and Verbs had most of the responses as neutral or negative.

There are some interesting observations when comparing what headline aspects people say they consider compared to our experimental results in Table 2. Prominence, Sentiment, and Simplicity were both positively judged in the survey and had relatively high significant correlations. However, some of the style features that had high correlations with social media popularity (e.g. Brevity, Nouns) had many neutral or negative responses in the survey. Conversely, Proximity which was associated with significantly lower social media popularity, had the highest proportion of positive responses. What could explain the differences in how news values and style are judged versus how they objectively influence popularity is that higher level concepts like Prominence or Proximity are more salient to readers than relatively technical concepts like nouns or adverbs. While style might not be perceived by readers to influence their choice of headlines (since headlines in news outlets should already be grammatical), it has a significant effect on social media popularity nonetheless. Furthermore, these differences might be due to a difference in the engagement with the headline – the survey asked about clicking on a headline (private engagement), whereas the correlations target explicit social behaviour such as likes or retweets. Overall, both the survey and the experimental results confirm that news values and style influence a headline's popularity.

[19]http://www.crowdflower.com

Conclusion

Headlines play a very important role in digital news spaces, especially on social media. In this paper we investigated two types of automatically extracted headline characteristics – news values and linguistic style – and their impact on article popularity on Twitter and Facebook. The news value of Prominence and style features of Brevity and Simplicity have proven to be especially well-correlated. Prominence, Proximity, and Sentiment were also judged as having a very positive effect when choosing to click on a headline by respondents in a crowdsourced survey. The methods outlined in this paper can be adopted by computational journalism researchers to further investigate the effect of headline features on popularity across different news sources and genres.

Our current and future work includes building a prediction model which uses news values and style features to predict the popularity of news values on Twitter and Facebook, and further developing the Prominence and Proximity features to take into account user location.

Acknowledgments

This work was supported by a Doctoral Training Grant from the Engineering and Physical Sciences Research Council, UK. Data collection and storage comply with EPSRC data management policies. The dataset is available at https://doi.org/10.5518/174.

References

Althaus, S. L.; Edy, J. A.; and Phalen, P. F. 2001. Using substitutes for full-text news stories in content analysis: Which text is best? *American Journal of Political Science* 45(3):pp. 707–723.

Baccianella, S.; Esuli, A.; and Sebastiani, F. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*.

Bandari, R.; Asur, S.; and Huberman, B. A. 2012. The pulse of news in social media: Forecasting popularity. In *ICWSM*.

Barr, C., ed. 2010. *The Yahoo! style guide: the ultimate sourcebook for writing, editing, and creating content for the digital world*. Macmillan.

Bednarek, M., and Caple, H. 2012. *News Discourse*. Continuum.

Bell, A. 1991. *The language of news media*. Blackwell Oxford.

Bell, A. 1995. News time. *Time & Society* 4(3):305–328.

Berger, J., and Milkman, K. L. 2012. What makes online content viral? *Journal of Marketing Research* 49(2):192–205.

Biber, D. 1991. *Variation across speech and writing*. Cambridge University Press.

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

Brône, G., and Coulson, S. 2010. Processing deliberate ambiguity in newspaper headlines: Double grounding. *Discourse Processes* 47(3):212–236.

Caple, H., and Bednarek, M. 2013. Delving into the discourse: Approaches to news values in journalism studies and beyond. *Reuters Institute for the Study of Journalism*.

Chovanec, J. 2014. *Pragmatics of Tense and Time in News: From canonical headlines to online news texts*. Pragmatics & Beyond New Series. John Benjamins Publishing Company.

Clarkson, P., and Rosenfeld, R. 1997. Statistical language modeling using the CMU-Cambridge toolkit. In *Eurospeech*.

Conley, D., and Lamble, S. 2006. *The Daily Miracle: An Introduction to Journalism*. Oxford University Press.

Cotter, C. 2010. *News talk: Investigating the language of journalism*. Cambridge University Press.

De Vries, L.; Gensler, S.; and Leeflang, P. S. 2012. Popularity of brand posts on brand fan pages: an investigation of the effects of social media marketing. *Journal of Interactive Marketing* 26(2):83–91.

Dor, D. 2003. On newspaper headlines as relevance optimizers. *Journal of Pragmatics* 35(5):695–721.

Feng, L.; Jansche, M.; Huenerfauth, M.; and Elhadad, N. 2010. A comparison of features for automatic readability assessment. In *Coling*.

Feng, S.; Kang, J. S.; Kuznetsova, P.; and Choi, Y. 2013. Connotation lexicon: A dash of sentiment beneath the surface meaning. In *ACL*.

Fowler, R. 1991. *Language in the News*. Routeldge.

Galtung, J., and Ruge, M. H. 1965. The structure of foreign news the presentation of the Congo, Cuba and Cyprus Crises in four Norwegian newspapers. *Journal of Peace Research* 2(1):64–90.

Gans, H. J. 1979. *Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek, and Time*. Northwestern University Press.

Gatti, L.; Özdal, G.; Guerini, M.; Stock, O.; and Strapparava, C. 2016. Automatic creation of flexible catchy headlines. In *Natural Language Processing meets Journalism Workshop*. IJCAI.

Graff, D.; Kong, J.; Chen, K.; and Maeda, K. 2003. English Gigaword. *Linguistic Data Consortium*.

Harcup, T., and O'Neill, D. 2001. What is news? Galtung and Ruge revisited. *Journalism Studies* 2(2):261–280.

Johnson-Cartee, K. S. 2005. *News narratives and news framing: Constructing political reality*. Rowman & Littlefield Publishers.

Kate, R. J.; Luo, X.; Patwardhan, S.; Franz, M.; Florian, R.; Mooney, R. J.; Roukos, S.; and Welty, C. 2010. Learning to predict readability using diverse linguistic features. In *Coling*.

Kepplinger, H. M., and Ehmig, S. C. 2006. Predicting news decisions. An empirical test of the two-component theory of news selection. *Communications* 31(1):25–43.

Klein, D., and Manning, C. D. 2003. Accurate unlexicalized parsing. In *ACL*.

Kucuktunc, O.; Cambazoglu, B. B.; Weber, I.; and Ferhatosmanoglu, H. 2012. A large-scale sentiment analysis for Yahoo! answers. In *WSDM*.

Lakkaraju, H.; McAuley, J. J.; and Leskovec, J. 2013. What's in a Name? Understanding the Interplay between Titles, Content, and Communities in Social Media. In *ICWSM*.

Leskovec, J.; Backstrom, L.; and Kleinberg, J. 2009. Meme-tracking and the dynamics of the news cycle. In *SIGKDD*.

Louis, A., and Nenkova, A. 2013. What Makes Writing Great? First Experiments on Article Quality Prediction in the Science Journalism Domain. *TACL*.

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

Piotrkowicz, A.; Dimitrova, V. G.; and Markert, K. 2017. Automatic extraction of news values from headline text. In *Proceedings of the EACL 2017 Student Research Workshop*.

Pitler, E., and Nenkova, A. 2008. Revisiting readability: A unified framework for predicting text quality. In *EMNLP*.

Quirk, R.; Greenbaum, S.; Leech, G.; and Svartvik, J. 1985. *A comprehensive grammar of the English language*. Longman.

Recasens, M.; Danescu-Niculescu-Mizil, C.; and Jurafsky, D. 2013. Linguistic models for analyzing and detecting biased language. In *ACL*.

Reis, J.; Benevenuto, F.; Olmo, P.; Prates, R.; Kwak, H.; and An, J. 2015. Breaking the news: First impressions matter on online news. In *ICWSM*.

Szymanski, T.; Orellana-Rodriguez, C.; and Keane, M. T. 2016. Helping news editors write better headlines: A recommender to improve the keyword contents and shareability of news headlines. In *Natural Language Processing meets Journalism Workshop*. IJCAI.

Tan, C.; Lee, L.; and Pang, B. 2014. The effect of wording on message propagation: Topic-and author-controlled natural experiments on Twitter. In *ACL*.

Toutanova, K.; Klein, D.; Manning, C. D.; and Singer, Y. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *NAACL-HLT*.

Tumasjan, A.; Sprenger, T.; Sandner, P.; and Welpe, I. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM* 178–185.

Vlachos, M.; Meek, C.; Vagena, Z.; and Gunopulos, D. 2004. Identifying similarities, periodicities and bursts for online search queries. In *SIGMOD*.