

The CALA 2020 Proceedings Paper 13 - 1

Text, Context, Entextualization, Paper 1

An Analysis of Indian English News Headlines

Nia Kurniasih

Department of Humanistic Studies, Indian Institute of Technology, BHU, Varanasi, India

Iis Kurnia Nurhayati

Department of Humanistic Studies, Indian Institute of Technology, BHU, Varanasi, India

Puji Audina Lestari

Department of Computer Science and Engineering, Indian Institute of Technology, BHU, Varanasi, India

An Analysis of Indian English News Headlines

Samapika Roy^a

^a*Department of Humanistic Studies, Indian Institute of Technology, BHU, Varanasi, India*

Sukhada^a

^a*Department of Humanistic Studies, Indian Institute of Technology, BHU, Varanasi, India*

Anil Kr. Singh^a

^a*Department of Computer Science and Engineering, Indian Institute of Technology, BHU, Varanasi, India*

Abstract

News Headlines (NHs) are of the most creative uses of natural languages in a media text. An NH is the frontline of a news article. Specific characteristics make NHs stand out: for instance, article omission, use of active verbs, dropping the copula to save space and to attract the reader's attention to the most significant words, etc. Some research has been done on linguistic analysis of British English NH, Hindi-Urdu NHs, but hardly any work has been conducted on IndENH.

This paper attempts to analyze Indian English newspaper headlines (IndENH), and aims to contribute to the accuracy of News Headline parsing. This study determines the linguistic features of the IndENH, to improve the quality of the parsed output of NHs. This paper covers sentence construction, tense, punctuation marks, metaphors, etc. for linguistic analysis.

Keywords: Indian English news headlines, Parsing, standard English, linguistic analysis.

Introduction

News stories have three prominent styles: Inverted pyramid style, hourglass style and chronological style (Seiha 2013). The news stories contain a headline, a lead, some quotes, important details and less important details. The most prominent amongst these is the headline. Some people act more ritualistically towards newspapers than others, i.e. some people read the entire newspaper where some are very specific in their choices of sections. The aim of an NH is to introduce the story to readers. NHs attempt to provide a very brief summary of news articles. These short summaries are designed carefully to grab readers' attention (BBC 2003). Some headlines encapsulate where some hint at the gist of the story that follows. Sometimes, tension emerges between the space limit of a news column and the space required for an NH to appear as a standard sentence. Due to several such motives, the editors are left with the only choices that render the headlines incomplete and often ungrammatical. Thus, headlines are a challenge for Natural Language Processing (NLP) tasks such as parsing, information extraction, machine translation, and so forth.

The motive behind this study is to identify the problems in NH parsing and improve the accuracy of NH parsing so that the NLP tools that use parsing produce better results.

Literature Review

Headlines have grammar which differs to that of ordinary sentences (Yoneoka 2002). From an NLP perspective, headlines pose an engineering challenge. They include linguistic aspects such as the unusual use of tenses (Chovanec 2014) and deliberate ambiguity (Brône and Coulson 2010). NLP researchers have focused on headlines, including headline generation (Banko et al. 2002) and translating NHs (Ono 2016). Some intertextual linguistic analysis (Fairclough 1995), focusing on lexical feature variations between Time Supplement NHs and The New York Times NHs (Shie 2010) has been conducted. A critical discourse analysis of online NHs has been conducted by Daria Lombardi to understand how online NHs represent the Stoneman Douglas High School shooting. Several other works such as the analysis of how attitudes are displayed in news reporting (Nordlund 2003; Pajunen 2008), a Linguistic and stylistic analysis of newspaper reportage (Agu 2015), tense in NHs (Hameed 2008) are further contributions. All of these above-mentioned works are sublime, yet we attempt to contribute to these through this study.

Data Collection

We created an IndENH corpus from print and digital media, from the top three Indian newspapers as per their circulation (Audit Bureau of Circulations). We analyzed *The Hindu* (TH), *The Times of India* (TOI) and *Hindustan Times* (HT) for this study. We included 1500 NHs taking 500 general titles from each newspaper. The three different newspapers facilitated the study of the structures of NHs, to confirm whether the discussion are newspaper specific or follow the same pattern.

Methodology

After data collection and sanitization i.e., removing the repeatedly occurring headings like section, cartoon, name of places, etc., we parsed these NHs using open-source parsers such as Stanford Parser and Allennlp Parser. We observed that the NHs are mostly incorrectly parsed. The types of mistakes parsers frequently make are: changing nouns into verbs, adverbs, proper nouns, present tense singular verbs as plural nouns and adverbs, and present tense plural verbs as adjectives; adjectives as verbs. Thus, it becomes necessary to analyze NHs linguistically to determine the problems and to correct these so as to increase the accuracy of parser output.

There is a necessity to compare the NHs with standard English, to understand the difference between the two. We studied the data through a comparative method, where we attempted to understand the structure and language use of NHs.

Linguistic Analysis

NHs are significantly different to grammatically correct standard sentences. Headlines have distinctive syntactic properties, which render them a grammatical oddity (Fairclough 1995). We have attempted to locate these distinctive linguistic properties in IndENHs. We found that all the IndENHs show similar linguistic features for all the different categories.

Types of News Headlines

In IndENHs, there are two types of NHs: Interrogative and Declarative.

Interrogative Headlines

Grammatical interrogative forms are verb forms/sentences/clause types used in the expression of questions (Crystal 2008). In NHs, we found ‘Simple Interrogatives’ and ‘Echo Questions.’

Simple Interrogatives: Simple interrogatives are the regular interrogative constructions, which ask a direct question.

TH: Doctors’ protest: Will govt. give in on contentious provisions of KPME Bill?

HT: Will Maharashtra Rera’s SRO filter benefit homebuyers eventually?

TOI: What happens to Rishabh Pant now?

Echo questions: Echo questions are pure statements, and do not involve wh-movement. They repeat someone’s statement with an interrogative tone marked via a question mark in NHs.

TH: Ranbir Kapoor plays a DJ in Brahmastra? An insider spills the beans.

HT: Spielberg’s stand cost Michael Douglas Cannes Glory?

TOI: Maharashtra to bail out 11,000 staffers with fake caste certificates

Non-Interrogatives: Headlines starting with a wh-word need not necessarily be an interrogative question. Such NHs are like situation starters whose answer is made available in the article if the reader chooses to know.

TH: When a crisis led to a new political bonding

HT: What does the Jewar airport’s revival mean

TOI: What exactly is Bhilwara’s strategy that cut infections

Declarative Headlines

Declarative headlines are the dominant and default headlines, mainly as the NHs state facts or report something that has happened. In the NHs, journalists inform readers of the events were mostly in the form of statements. For instance,

TH: Medical services in Mysuru likely to be hit today

HT: China isolated on Jammu and Kashmir in informal UNSC talks

TOI: Kuldeep Singh Rathore named as chief of Himachal Congress

Until now, examples emerge from all three newspapers. Hereinafter, we will present examples only from *The Hindu*.

Compound Noun

NHs present a very creative usage of noun compounding. There is a difference between the compound nouns found in literature and those found in NHs. The compound nouns in NHs are contextual, i.e. they are created as per the context just as shown below:

CM criticises Yeddyurappa for *corruption accusation*

Dropping of Subject NP

Dropping the subject NP from the headlines where it is not crucial to editors, can be found in NHs. In the given example, someone or something is turning the used oil into fresh fuel, but the subject NP is deliberately dropped, thus, creating a lack of transparency:

Turning used oil into fresh fuel

Noun Phrases Containing Only Cardinals

The cardinals such as 16, 9, and five are followed by equivalent noun subjects syntactically, but in NHs they are dropped. These cardinals may sometimes be directly followed by adjectives, past participles, and so forth, omitting the head noun of the subject NP. This noun omission leads to a lack of transparency.

1. Numerical + Adjective: 16 dead as boat capsizes in Krishna
2. Cardinal + Past Participle: *Over 350 screened at eye camp*

Auxiliary Dropping

Auxiliary verbs are subordinate verbs that make distinctions in tense, mood, aspect, voice (Crystal 2008). Dropping an auxiliary is the most common way to save space in NHs. Mostly, the main verb carries the tense, thus informing the readers of the status of the event. For example, one may add the auxiliary 'was' in the NH '*Dengue awareness rally held*' as '*Dengue awareness rally was held*' to arrive at a standard English sentence.

Tense

Tense marks the time at which the action denoted by the verb occurred (Crystal 2008). The NHs use tense to mark the time reference of the event.

Present Tense: The present tense is the predominantly used tense in NHs. It is used not only for ongoing events but also for reporting a past event that happened within a span of 24-48 hours i.e. is hard news whose impact lasts for 24-48 hours (Reah 1998). 'News broadcasts and newspapers are designed to make one think that news stories are happening NOW' (BBC 2003). In the present tense, two forms are found:

Simple Present: Simple present is referred to as 'Historic Present.' This tense is used even when the event has already passed but the editors want to keep the readers involved in a present time frame because readers feel more interested in reading about an event that is happening now. A past event in the past tense is less impactful. Ex.

Heritage steam engine runs 2 km without driver

Present Progressive: Present progressive or present continuous tense shows an ongoing event. This is also used when the event is longer than a day and is ongoing.

*Pinarayi **protecting** encroachers*

Past Tense: Past tense refers to an event that has already happened. Verbs with past tense are the second most used verbs in NHs. The main verbs of the NHs in the past tense have two forms:

Simple Past: In NH, the simple past is used to mark the past time reference. The auxiliary denoting past is mostly dropped.

Launch of double bedroom houses delayed

Past participle: Editors portray the past perfect tense through past participle.

*Sandalwood trees **stolen** from C.V. Raman's home*

Future Tense: Future tense is not considered as a tense per se (Comrie 1985), yet NHs still use it to mark the upcoming events. NHs usually employ verbs in their infinitive forms to denote a forthcoming event (BBC 2003). The future tense is the least used tense in the NHs amongst the three.

*India to **invite** Pakistan PM Imran Khan for SCO meet*

Different Functions of Punctuations

Comma: The comma “,” is used to mark clause boundaries, but in NHs, it has different functions based on their positions.

Speech-Speaker: Here, a comma is used to draw a clause boundary between the speaker and the speech with the help of a verb. The speaker can be a person or an organization. This construction is used to show the study and its report as well.

Need an NIA unit in Mangaluru, says BSY

This NH could be rewritten as ‘BSY said there is a need for an NIA unit in Mangaluru’

Conjoining two clauses: Two clauses representing two incidents are conjoined by a comma.

Teenager killed, friend injured in accident

Conjunction: The comma also functions as a conjunction.

Call for unity between Kapus, SCs

This NH can be rewritten as ‘Call for unity between Kapus and SCs.’

Cause-effect: In this case, the function of the comma is to link two clauses, one of which states the cause and the other the effect. Ex.

Bhalswa landfill fires, smog have residents in chokehold

The NH can be rewritten, interpreted as ‘smog has residents in a chokehold as Bhalswa landfill is fired.’

Colon: The colon “:” is used for listing, to indicate a sub-title or subdivision of a topic, but in NHs, it performs different functions in a different position.

Speech-speaker: NHs having two clauses conjoined by colon stating a speech-speaker relationship between the clauses. Ex.

Beating up of teenager: mother to launch indefinite fast

The given example can be rewritten as ‘mother is going to launch indefinite fast for beating up of teenager.’

Topic-Information: In this case, one clause denotes the topic and the other denotes relevant information (progress/regress).

Businessman murder case: No arrests yet

The first part of the NH, i.e. 'Businessman murder case,' states the 'topic' and the second part 'No arrests yet' informs about the progress.

Semi-colon: The semi-colon “;” is used for listing, however, NHs also use it to join together two clauses, each stating different things about a scenario; one states an incident and the other states the result. A comma is also used when the editor has more than one outcome to list.

Incident-result relation:

UP police back in encounter mode; 1 killed, 5 injured

Miscellaneous

Graphological devices play a significant role in ephemeral texts such as newspaper headlines. Apart from the interesting usage of punctuation marks, few other stylistic techniques are used while creating an NH. Some of these are discussed here.

Short Forms: Short forms shorten words by clipping. Ex:

Maharashtra govt. ads draw flak

Abbreviations: Abbreviations are shortened forms of phrases and are the most useful technique with which to minimize space crises.

South China Sea, Rohingya on ASEAN map

Code-mixing: Code-mixing refers to the melding of linguistic elements from two (or more) languages or dialects. In the given example, the Urdu word 'talaq' meaning divorce is used in an English NH.

AMU professor's wife alleges talaq on WhatsApp

Capitalization: English capitalizes the first letter of a sentence, but in NHs, the entire word or the headline is capitalized mostly to get the reader's attention.

Ear SHOT ashok rajagopalan

TWO DEAD DURING JALLIKATTU IN TN

Metaphor: NHs are often crafted metaphorically, to add some color or likeness between the story and the real-world event.

It is raining groundnuts

Ambiguity: NHs introduce ambiguity deliberately so as not to give away any clear idea, but rather to draw the reader's attention to the NHs and to compel them to read the articles.

Medical services in Mysuru likely to be hit today

'To be hit' can mean either 'will be successful,' or 'will be affected,' or 'will start from today.'

Personification: NHs creatively attribute human characteristics to non-humans.

Facebook says technical error caused vulgar translation of Xi Jinping's name

Here, the app 'Facebook' is personified for the people running the app.

Puns with Homophones: Homophones are words with similar sounds but different meanings.

Hiding in 'Plane' sight.

The idiom 'hiding in plain sight' has been twisted to intend the pun it generates through replacement with 'hiding in 'Plane' sight.' 'Hiding in plain sight' means to be unnoticeable. By hiding in 'Plane' sight the authors intended to suggest that it is hard to find an airplane in the picture of the article.

Article Omission: Article (A, An, The) omission is the most commonly used technique in NHs, deliberately used to save space or intentional concealment of specificity.

Biker dies in mishap

Scare-Quotes: In NHs, editors' use of quotation marks for a non-standard, irony.

Road map to regaining 'cleanest' tag

Quotes Without Speaker: An NH could sometimes be a mere quote. Editors intentionally drop the speaker in such NHs so as to generate curiosity and to compel the readers to read the article.

'Aggression, violence are a reality of the world we live in today'

Conclusion

This paper has deliberated upon the linguistic analysis of news headlines from three Indian English newspapers. A news editor's perception is different from that of grammarians or linguists, the lack of space and need to grab readers' attention leads to very creative usage of language in the form of headlines. The study found that all three newspapers follow the same linguistic patterns. From contextual compounding to using present tense for past events news editors take creative liberty to construct well crafted NHs. Dropping of phrases, nouns, pronouns, determiners and auxiliaries, using short forms, abbreviations along with usage of rhetorical devices like metaphor, personification, and so forth, render NHs attention-grabbing. One of the most fascinating points is the creative usage of punctuations, as editors render the punctuation marks as speaking more than the basic form. NHs appear to be becoming increasingly informal, a suitable exemplar for which will be the use of code-mixing and scare quotes. NHs use short forms and can also be found with ambiguity, personification and homophonic puns.

References

- Agu, Innocent Ejimofor. "A Linguistic-Stylistic Analysis of Newspaper Reportage. International Journal of Research in Humanities and Social Studies." Volume 2, Issue 8, (August 2015), PP 20-27. ISSN 2394-6296 (Online)
- Banko, Michele and Mittal, Vibhu and Witbrock, Michael. Headline Generation Based on Statistical Translation. (2002). 10.3115/1075218.1075259. BBC. 'THE LANGUAGE OF THE MEDIA.' (2003).
- Bell, A.D. "The language of news media." Wiley-Blackwell Publication. 1991.

- Bly, Bob. Some Types of Headlines. Philippine Content Development. (2011).
- Bonyadi, Alireza and Samuel, Moses. Headlines in Newspaper Editorials: A Contrastive Study. SAGE Open April-June 2013: 1–10. (2013).
- Brône, Geert and Coulson, Seana. Processing Deliberate Ambiguity in Newspaper Headlines: Double Grounding. *Discourse Processes - DISCOURSE PROCESS*. 47. 212-236. 10.1080/01638530902959919. (2010).
- Chovanec, Jan. 2003. "The Uses of the Present Tense in Headlines." *Theory and Practice in English Studies*, Vol. 1, ed. by Jan Chovanec, 83–92. Brno: Masaryk University.
- Chovanec, Jan. *Pragmatics of Tense and Time in News: From Canonical Headlines to Online News Texts*. (2014).
- Crystal, David. *The Encyclopedia of the English Language*. Cambridge: Cambridge University Press. 1988.
- Crystal, David. *A dictionary of Linguistics and Phonetics*. Cambridge: Cambridge University Press. 6th Ed. 2008.
- Crystal, David, and Derek Davy. *Investigating English Style*. London and Harlow: Longmans. 1969.
- Develotte, Christine and Rechniewski, Elizabeth. *Discourse analysis of newspaper headlines: a methodological framework for research into national representations*. (2001).
- Fairclough, Norman. *Critical Discourse Analysis: The Critical Study of Language*. Harlow: Longman. (1995).
- Hameed, Hind. *Tense in Headlines*. (2008).
- Hasan, Lect and Abu-Krooz, Hasan and Qasim, Obayes and Al-Azzawi, and Majid, Mohammed and Saadoon. *Code Switching and Code Mixing: A Sociolinguistic Study of Senegalese International Students in Iraqi Colleges*. (2019).
- Isani, Shaeda. *Of headlines and headlines: Towards distinctive linguistic and pragmatic genericity*. p. 81-102. (2011).
- Lombardi, Daria. *Critical Discourse Analysis of online News Headlines: A Case of the Stoneman Douglas High School Shooting*. (2018).
- Lodhi, Muhammad Arfan et al. *Textual and Rhetoric Analysis of News Headlines of Urdu and English Newspapers*. *International Journal of English Linguistics*; Vol. 9, No. 1. (2018).; Canadian Center of Science and Education. ISSN 1923-869X E-ISSN 1923-8703.
- Moe, Sandar A Brief Study on the Language of Newspaper Headlines Used in "The New Light of Myanmar." *Hinthada University Research Journal*. Vol. 5, No. 1. (2014).
- Napu, Novriyanto. *English And Indonesian Newspaper Headlines: A Comparative Study Of Lexical Features*. *European Journal of Literature, Language and Linguistics Studies*. ISSN:2559 - 7914. Vol. 2. Issue 4. (2014).
- Nir, Raphael. *A discourse analysis of news headlines*. *Hebrew Linguistics* 37. (1993).
- Nordlund, Marie. *Linguistic Manipulation: An Analysis of How Attitudes are Displayed in News Reporting*. (2003).
- Ono, Kenji. *Translation of news headlines*. (2003).
- Pajunen, Juhani. "Linguistic Analysis of Newspaper Discourse in Theory and Practice." (2008). Pape, Susan and Featherstone, Sue. *Newspaper Journalism: A Practical Introduction*. SAGE Publication. (2005).
- Shie, J. *Lexical feature variations between New York Times and Times Supplement News headlines*. *Concentric: Studies in Linguistics*, 36(1), 79-103. (2010).
- Reah, Danuta. "The language of newspapers." (1998).
- Reah, D. *The language of newspapers*. London, England: Routledge. 2002.
- Van Dijk, Teun A. *News as Discourse*. Hillsdale, Hove and London: Lawrence Erlbaum Associates. (1988).
- Yoneoka, Judi. *English in Asia*. (2002).

Yergaliyev, Kuanysh Sovetuly. Newspaper Headlines as a Marker of a Language Picture of the World. Indian Journal of Science and Technology, Vol 9(26). (2016). ISSN (Online) : 0974-5645.

<http://www.thehindu.com/archive/print/2017/11/13/>.

<http://www.thehindu.com/archive/print/2017/11/14/>.

<https://epaper.hindustantimes.com/Home/ArticleView>. dated- 06.12.19 to 09.12.19

<https://epaper.timesgroup.com/Olive/ODN/TimesOfIndia/Default.aspx>.