“Busted!” “…And He's Got a Helluva Explanation for Why”

The Language Use in U.S. Based Online Celebrity Journalism

Hanna-Mari Rämä

Supervisor: Roman Horbyk
Södertörn University | School of Social Sciences
Master Thesis 15 credits
Journalism | Spring 2018
International Master’s Programme In Journalism
Abstract

This thesis investigates which linguistic phenomena and features are shown by online celebrity news articles in order to identify how celebrity is constructed linguistically and what kind of language journalists use in their work when reporting on celebrity. Two American based celebrity news websites, Radar Online and TMZ were chosen for this investigation. Additionally, the thesis compares in what way the style and language varies between these two websites. Both qualitative and quantitative analyses using critical linguistics, rhetorical analysis and corpus linguistics have been conducted on 180 celebrity news articles in total. After the analyses, the study was able to identify that the most frequently used linguistic phenomena in the articles are within the area of non-standard vocabulary and lexicon forms, that is, unconventional spelling, colloquialisms, slang and taboo language. Furthermore, the analyses showed that even though the celebrity websites are thematically similar and have similar content, their writing style and language usage deviate from each other. In addition, the study indicates that celebrity is constructed linguistically by using particular kind of vocabulary, simplistic language and spoken language features bringing conversational elements to the writing. Celebrity is also constructed by the usage of the word star, rather than the word celebrity itself.

Keywords: celebrity culture, celebrity news, celebritification, tabloidization, corpus linguistics, rhetorical analysis, critical linguistics, Radar Online, TMZ
# Table of Contents

1 Introduction
   1.1 Background 1
   1.2 Aim and Research Questions 1
   1.3 Structure of the Study 2
   1.4 Limitations of the Study 3
   1.5 The Selected Celebrity News Outlets 3
      1.5.1 Radar Online 3
      1.5.2 TMZ 4

2 Theoretical Framework 4
   2.1 Celebrity and Celebrity Culture 4
   2.2 Tabloids and Tabloidization 6

3 Previous Research 8
   3.1 Studying Celebrity and Celebrity News 8
   3.2 Celebrity News as a Genre 9

4 Methodology and Empirical Data 10
   4.1 Design of the Present Study 10
   4.2 Linguistic Phenomena 12
   4.3 Critical Linguistics 12
   4.4 Rhetorical Analysis 14
   4.5 Corpus Linguistics 17
      4.5.1 Compiling the Corpora 18
   4.6 Reliability and Validity 21

5 Results and Analysis 22
   5.1 Identifying the Different Types of Linguistic Phenomena 22
   5.2. The Characteristics of the Celebrity News Articles 23
   5.3. Critical Linguistics and Rhetorical Style 27
   5.4 Corpus-Based Analysis 31

6 Discussion 43
   6.1 Results of this Study 43
   6.2 Celebrity Journalism – Hybrid Journalism? 46

7 Conclusions 48

8 References 51

9 Appendices 54
1 Introduction

1.1 Background

Some time ago, celebrity news coverage was limited mostly to tabloids and the like. Nowadays, celebrity is a fundamental component of the news in all media platforms and it has spread throughout the whole media spectrum: celebrity coverage has become omnipresent to the extent that it constitutes a new normality in the contemporary media world (Dubied & Hanitzsch, 2014). What has made celebrity news so popular, and why do people read these “guilty pleasure” (Dubied & Hanitzsch, 2014) news stories?

In order to set grounds for this study, celebrity news must be defined and discussed. According to Turner (2014), the term “celebrity news” itself is already problematic, as it “looks like a blatant pitch for legitimacy” (p. 144), and it should be dealt with some level of skepticism. Generally speaking, celebrity news is a widely used industry descriptor for commenting and reporting on celebrity. It is probably the only area of journalism where readers and consumers can accept gossip as news without hesitation (Turner, 2014).

Celebrity news is a special news genre in which reporting on scandals, gossip, rumors and speculations is not judged by in terms of accuracy, but in terms of its demonstration of the commentator or reporter’s access to stories, that would be unavailable otherwise, even if the stories turned out to be false. In reporting and reading celebrity news, there is the appeal of something that looks like an insider perspective, and audiences enjoy uncovering the secrets of celebrities and public figures. It could be argued that the sole objective of celebrity news is to entertain its audience. Regardless of its wide use and distribution, and nowadays strong connections to all media platforms, celebrity journalism for some is still not a completely respectable branch of journalism (Turner, 2014).

Celebrity news is also seen as “soft” news; soft news topic comprehend entertainment, lifestyle and celebrity news, whilst by “hard” news is generally meant topics that are timely and consequential, such as politics, business and international affairs (Reuters Institute, 2016).

1.2 Aim and Research Questions

This study focuses on celebrity news coverage as a genre from a linguistic perspective. The study investigates the language of two celebrity news outlets that are U.S. based and written in English, Radar Online and TMZ. The U.S. based celebrity media is chosen for the investigation as the American celebrity culture formulates the biggest entertainment industry
in the Western countries (Van den Bulck & Claessens, 2014). The study consists of qualitative and quantitative research approaches, using critical linguistics, rhetorical analysis and corpus linguistics. Additionally, the study seeks to identify in the celebrity news outlets the linguistic phenomena and stylistic features that are typical for celebrity news coverage. The investigated celebrity news outlets and applied research methods are introduced and discussed in more detail in the following sections.

The study is situated in journalism and media studies, as it seeks to investigate the language of a specific type of medium and how the language is constructed and employed in that medium. Celebrity news is a frequently discussed journalistic genre in the context of tabloidization, and therefore an interest for journalism studies (Van den Bulck & Claessens, 2014). Since the celebrity news coverage is viewed as a genre in this study, linguistic devices are required for the analysis, as language structures can be seen as fundamental part of the social contexts and functions of a text (Bawarshi & Reiff, 2010).

The aim of this study is to investigate, how journalists use language when reporting on celebrity. The study is concerned with, how celebrity is constructed linguistically; therefore, this investigation has the potential to highlight celebrity journalism as a specific genre, and how a specific type of language is required for mediated representations of celebrity. The research questions are:

**RQ1**: What kinds of linguistic phenomena, features and structures can be identified in the celebrity news coverage?

**RQ2**: What are the most frequently occurring linguistic features and phenomena of the celebrity news coverage?

**RQ3**: What kinds of differences are there between the selected celebrity news outlets?

### 1.3 Structure of the Study

Firstly, the selected celebrity news outlets, *Radar Online* and *TMZ* are introduced briefly. Next, the theoretical concepts that are relevant for the investigation are discussed. The theoretical concepts are: celebrity, celebrity culture and celebritification, followed by tabloids and tabloidization. Then, the previous research that has been conducted in the field of celebrity journalism is touched upon briefly as well.

Secondly, the empirical data and methodology are introduced and discussed. These include both qualitative and quantitative research approaches, applying identification of
linguistics phenomena and devices, critical linguistics, rhetorical analysis and corpus linguistics.

Thirdly, the results of the analysis, followed by a discussion and a conclusion are presented and pondered.

1.4 Limitations of the Study
Certain limitations of the study have to be noted though. First of all, this study focuses only on U.S. based celebrity news outlets, excluding other big popular culture markets, such as United Kingdom, continental Europe and Australia. Thus, the perspective is very Western orientated.

Second of all, only two celebrity news outlets are chosen for the analysis. While there would have been many more thematically similar news outlets, only two were included for this investigation, given the scope and time limitations of the study.

Third of all, both of these celebrity news outlets are online based, thus printed celebrity news coverage is not included in the analysis.

And finally, the timeframe that is examined in this study covers only one month; therefore, the analysis concentrates solely on very topical (and recent) celebrity news coverage. On the other hand, it has to be noted though, that the study is not a longitudinal study. Consequently, what is examined and analyzed is the present and up-to-date language of celebrity news coverage.

1.5 The Selected Celebrity News Outlets
1.5.1 Radar Online
*Radar Online* is an American gossip and celebrity website that was first published in a printed version, and since 2009 it has been solely an online publication. It is owned by the news organization *American Media, Inc.*, according to which the website *radaronline.com* has nearly ten million visitors a month and nearly 70 million monthly page views (American Media Inc., n.d).

They advertise themselves as “the most influential, trusted celebrity & pop culture news brand on the planet” (American Media Inc., n.d). *Radar Online* creates over 50 exclusive pieces of content every day that take over the entertainment and media landscape 24 hours a day (American Media Inc., n.d). *Radar Online* also pays for “juicy information” and tips for stories (Radar Online, n.d.).
1.5.2 TMZ

A Hollywood and entertainment centered website, TMZ, was launched in 2005 and it is owned by America Online Inc. (AOL) and Telepictures Productions (America Online Inc., n.d.). The website concentrates on celebrity and Hollywood gossip, and it includes interviews, videos and photos of celebrities, as well as news on movies, TV shows and games. The initialism “TMZ” stands for “thirty-mile zone”, which refers to the Hollywood entertainment industry (Mediaweek, 2005).

According to McNamara (2011), TMZ has been an important factor in the popularization of the genre of celebrity gossip, and it has contributed remarkably to entertainment news coverage. Famous celebrity scandals and gossip that TMZ reportedly has broken through first are, for example, Britney Spears’s divorce in 2007, and the death of Michael Jackson in 2009 (McNamara 2011).

2 Theoretical Framework

2.1 Celebrity and Celebrity Culture

Celebrity culture has a widespread presence in the modern society, and there is no sign that it has reached its limits: besides popular culture and entertainment industries, celebrity culture has spread to politics, business world, sports, literary publishing, and so forth. Celebrity culture shapes the production and consumption of media content, as well as the social values through which the world is experienced (Turner, 2010a; Holmes, S. & Redmond, S., 2006).

The modern meaning of the term celebrity descends from the fall of the gods and the rise of the democratic, secular societies and governments. The origin of the word celebrity derives from a Latin word celebrem, which has connotations with “fame” and “being thronged” (Rojek, 2001, p. 9). But what constitutes a celebrity? Daniel Boorstin coined one of the most famous aphorisms about celebrity in 1971 (as cited in Turner, 2013, p. 4): “the celebrity is a person who is well-known for their well-knownness”.

Analyzing celebrity, celebrity culture and celebrities has been one of the growing industries in humanities and social sciences over the last decades. In fact, in cultural and media studies, celebrity has become part of the heartland for the study of popular culture (Turner, 2010a).

Celebrity can be described in various ways: as a representation, a discourse, an industry and a cultural formation (Turner, 2010a). Turner (2013) defines celebrity in the following manner:
celebrity is a genre of representation and a discursive effect; it is a commodity traded by the promotions, publicity, and media industries that produce these representations and their effects; and it is a cultural formation that has a social function we can better understand (p. 10).

According to Rojek (2001), mass-media representation is the most important feature of the formation of celebrity culture. Celebrities are cultural fabrications, and the rise of celebrity as a public preoccupation is the result of three important historical processes: the democratization of society, the decline in organized religion and the commodification of everyday life. In other words, as modern society has developed, celebrities have to a great extent replaced monarch and religion as new symbols of recognition and belonging. Celebrity culture thus provides an essential function of integration in secular society (Rojek, 2001).

Types of celebrities can be solved into different categories. Ascribed celebrity usually follows bloodline and is inherited; royalties are the best example of this category, but also children of famous people tend to become famous because of their family. Achieved celebrity derives from the professional achievements and accomplishments of the career of the celebrity, for instance athletes, actors, musicians, and the like, who have become famous for their skills and talent, are examples of achieved celebrities. However, achieved celebrity can also become attributed celebrity, which is a situation when the personal life of a celebrity is more of an interest than his or her professional life. The mass media and sensationalism are the main reasons for becoming attributed celebrity, as sensationalism aims to arouse public interest with the object of “spicing up” public attention (Rojek, 2001). Women tend to become attributed celebrities more likely than men (Holmes, S. & Redmond, S., 2006).

Closely related to attributed celebrity is celetoid, a concept introduced by Rojek (2001). These are individuals who have no particular talent, nor career objectives in the entertainment industry, other than to be famous and to be on the spotlight. In most cases, they will be famous for a little while, but they will be forgotten just as fast as they got famous. Examples of celetoids are mistresses of public figures, one-hit wonders, have-a-go-heroes, social media personalities and reality TV stars and contestants. They are “ordinary” people who push themselves through celebrity culture and into the public eye (Turner, 2006; Turner, 2010b).

Driessens (2012) calls the process by which these “ordinary” people are transformed into celebrities by the term celebrification, the making of a celebrity. Social media platforms, do-it-yourself-websites and reality TV shows are examples of this process. The expansion of reality TV formats, for example, have made the celebrity production and
celebrification shift from the “elite” to “ordinary” people (that is, no special talent is needed anymore to become famous, instead appearing in the media and simply being famous is sufficient enough), and the celebrity production has become a manufacturing process, creating replaceable celebrity commodities (Turner, 2006).

Driessens (2012) adds that celebrification also involves commodification. Celebrities are labor, and as such they are manufactured by the celebrity industry, producing and helping to sell other commodities. This is natural, as Turner (2013) explains, because celebrities are developed to make money. They are used to market anything from films and magazines to television programs and even media events and projects; they are “property”. Therefore, celebrity culture also represents the face of capitalism (Driessens, 2012).

Celebrity and celebrity culture are significant for this study since the material that is analyzed focuses on celebrity news. The theoretical concepts of celebrity, celebrity culture and celebrification are applied in the study in order to identify how celebrity is constructed and mediated linguistically. In addition, understanding the phenomena may help to uncover, why certain linguistic choices are made when reporting on celebrity.

2.2 Tabloids and Tabloidization

Celebrification can occur in many journalism genres, but most commonly in tabloids. Tabloid is a form that is marked by two important features: it offers little attention to politics, economics and society, and much attention to sports, scandals, celebrity and entertainment. Tabloid newspapers and magazines typically include sensational and personalized news style, blurring boundaries between private and public, politics and entertainment, celebrity-orientated and sexualized news agenda, and populist and partisan political interventions. In addition, tabloids are generally known for aggressive journalistic methods as well, such as cheque-book journalism (paying for stories or tips that lead to juicy stories) and paparazzi coverage. Examples of tabloid newspapers are the British newspapers the Sun and the Mirror, and the American magazine People (Johansson, 2007; Turner, 2014).

Tabloids usually consist of a predominant use of visual images and dramatization over textual elements, as paparazzi images are a central aspect of tabloid newspapers and magazines (Johansson, 2007; McNamara, 2011). Paparazzi photographers are a staple in the entertainment journalism, generating evidence of celebrities in newsworthy contexts, such as celebrity scandals, romance, candid photographs and the like. As the digital era has created new media outlets that are online, the paparazzi industry is booming and the output of celebrity photographs has increased as well, resulting in paparazzi websites that are
dedicated for celebrity photographs only. Thus the growing paparazzi industry has become a media in its own right (McNamara, 2011).

According to Johansson (2007), there are three primary reasons why people find tabloids attractive to read: 1) they are fun, they provide entertainment, and they operate as an opposite for depressing news, such as crime and disasters, 2) reading them can function as an attack on social privilege (especially reading celebrity coverage) and finally, 3) reading tabloids is linked to experiences of community and belonging. Johansson (2007) states that the interest in tabloid celebrity stories can be seen as a commentary on certain social tensions and power struggles in the society, in which they operate in the mode of “where pleasure taken in venomous or revelatory treatment of celebrities must be seen in the light of what they symbolise in relation to the lives of readers” (p. 148). The bashing of a celebrity in tabloids can offer a momentary experience of power and control to readers. In addition, reading tabloids can be seen as a social activity, strengthening social bonds and sharing experiences in the social sphere (Johansson, 2007).

Tabloidization involves a variety of claims about a change in the media production and content, and the beginning of its contemporary form is usually located to 1970’s, although the long history of the tabloid press can be traced back to the 19th century (Johansson, 2007; Turner, 2010b). Tabloidization, or rather tabloid culture, was originally “a tabloid term itself” (Turner, 2010b, p. 28), an expression of the media industry rather than an academic concept that was denoting a dumbing down media content and weakening the idealistic functions of mass media. Basically, while tabloidization originally was meant to describe a decline in journalism standards, nowadays the term is increasingly applied to all types of media content, for example talk shows, reality television, celebrity magazines, gross-out comedy, documentary, and so forth. Simply put, tabloidization can be defined as popularizing the news (Turner, 2010b).

Causes for tabloidization could be explained by technological changes, as new technologies (for example the great amount of cable and satellite channels and Internet sources) put journalists under more pressure than before to produce news quickly, which can lead to weaker quality and less time to check the facts; this can make the news seem more entertainment orientated. Another explanation for tabloidization could be in economic trends, in relation to internationalization and deregulation of media ownership and content. For example, “a well-documented international process of conglomeration is placing media organizations within larger corporate structures controlled by non-media interest, which,
while not necessarily leading to changes in media content, create greater potential for proprietors and commercial interests to influence editorial policy”, (Johansson, 2007, p. 33).

Tabloidization has been criticized of “dumbing down the news content and journalistic quality” and distracting audience’s attention from important issues (Dubied & Hanitzsch, 2014; Turner, 2014). On the other hand, in regards to tabloidization (as the tabloid culture typically evolves around celebrity content), the status of celebrity simplifies the determination of news value, as the level of fame of a person establishes its newsworthiness. In other words, while other news events may not attract readers to the same extent, celebrity often guarantees a high level of interest (Marshall, 2006).

It is important for this study to understand tabloidization and the tabloid format, as their language use, writing style and central topics generally differ from other types of news and media formats. While the celebrity news outlets that are analyzed in this study are online publications, and as such they represent new celebrity media formats, the language devices and stylistic features in them are still similar to the “old” printed tabloid forms (McNamara, 2011). Tabloids and tabloidization are thus applied in the analysis to identify what are the typical phenomena that can be seen as characteristic and representative for celebrity news coverage.

3 Previous Research

3.1 Studying Celebrity and Celebrity News

The study of celebrity news falls under a number of different disciplines. Communication and media studies have researched tabloidization and sensationalism in the news, as well as the readers of tabloid journalism (Johansson, 2007; Dubied & Hanitzsch, 2014). Furthermore, with new types of media formats arising, celebrity news has been studied from the aspects of paparazzi industry, as well as from the perspectives of production and consumption of celebrity news and gossip websites. In addition, audience reactions on celebrity news websites have been investigated as well (McNamara, 2011; Van den Bulck & Claessens, 2014), to give a few examples.

Cultural studies, on the other hand, have produced a rather large amount of studies on celebrity, although not necessarily so much about news and news production (Dubied & Hanitzsch, 2014). Turner (2010a) argues that the field of celebrity studies is populated with analyses of individual celebrities, either as media texts, or as pointers to broader cultural formations or political issues. Additionally, celebrity studies have produced new formations of theories and history of celebrity (Rojek in 2001), as well as studies of
celebrity production (Turner et al. in 2000), and studies of those who consume celebrity (Hill in 2007). What is more, other disciplines, such as psychology and sociology have shown interest in celebrities, and studied celebrity in terms of how young people perceive celebrity and fandom, to name but a few examples (Turner, 2010a).

According to Dubied & Hanitzsch (2014), when studying celebrity, authors do agree that celebrity is inseparably bound with modern mass media, but as the focus of their work is usually merely on celebrity culture, the journalism and communicative processes are mostly treated as contexts or givens.

3.2 Celebrity News as a Genre

From the perspective of journalism, however, celebrity coverage does not yet constitute an important area in the field of media and communication studies, as it tends to focus more on “hard news”, such as politics and current affairs. What is more, the normative and structural contexts of communication and media studies further contribute to the contentious appeal of celebrity news studies (Dubied & Hanitzsch, 2014).

Dubied & Hanitzsch (2014) argue that it is “easier to defend a study of conflict news and election coverage”, as the relevance of these types of news “is fairly undisputed in the public arena and scholarly world, while following celebrity often gets discredited as ‘guilty pleasure’” (p.140).

As a genre, celebrity news is both reveling and worthwhile on the one hand, and “at the heart of journalistic disputes on the other” (Dubied & Hanitzsch, 2014 p. 141). Celebrity news coverage is at the forefront of an explanation of the social function of news, while it also poses a number of challenges to journalism and its research. For example, celebrity news often transgresses boundaries between fact and fiction, bearing important questions for defining news and professional identities of journalists (Dubied & Hanitzsch, 2014).

This phenomenon is closely related to fake news; fake news means fabricated information mimicking news media content in form, but not in organizational process or intent. Fake news lacks news media’s editorial processes and norms for ensuring the credibility and accuracy of information (Cornell University, 2018). Celebrities are easy targets for fake news, as celebrity fake news stories are raunchy, and they spread fast (especially in social media), and fact checks, which could possibly limit the reach of celebrity fake news, cannot keep up (Funke, 2018). Also, celebrity news sets up challenges to media
laws and professional codes of conduct, alluding to tensions between public interest and protection of privacy (Dubied & Hanitzsch, 2014).

This study seeks to contribute to the field of journalism studies by investigating how celebrity is constructed linguistically, and how the language that is used when reporting on celebrity is employed and applied in the genre of celebrity journalism.

4 Methodology and Empirical Data
4.1 Design of the Present Study
This study comprises analyses of linguistic features of online celebrity news coverage. It is a qualitative study mixed with quantitative approaches, consisting of an investigation of two celebrity news outlets from the USA, as the U.S. celebrity culture represents the largest popular culture industry in the Western countries (Van den Bulck & Claessens, 2014). The research methods that are used in this study are critical linguistics, rhetorical analysis and corpus linguistics. In addition, identification of linguistic phenomena is also conducted.

For the analyses two celebrity news websites were chosen, Radar Online and TMZ. The websites only exist as online publications, and as such, they represent new celebrity media formats, as opposed to the “old” or “traditional” tabloids in forms of printed newspapers and magazines (McNamara, 2011), which is why these websites were chosen for the study. Another reason for choosing Radar Online and TMZ for the investigation is their popularity online and the large amount of news pieces and articles they produce on a daily basis. For example, Radar Online creates over 50 pieces per day (American Media Inc., n.d).

The timeframe that was followed in the study covers one month from 10th of March to 10th of April 2018. The timeframe has been selected in order to get as recent and topical celebrity news publications as possible. The amount of articles that were examined is 90 articles from each media outlet, adding up to 180 articles in total. The sample of 90 articles from each website was chosen in order to get as solid understanding of the style and language usage of the websites as possible, but within the limits of the study, the amount of the articles was drawn to 180 in total. What is more, 90 articles per website were seen sufficient for identifying the linguistic patterns and phenomena. The articles were chosen by using random sampling.

The randomization of the article selection was conducted by using MS Excel. Firstly, the number of the articles per day was counted and each article was given a number. These numbers corresponded to in what order the articles were published that given day. Secondly, the numbers of the articles were put in one column in MS Excel. Thirdly, the MS
Excel command RANDOM() was used in order to generate a random number between 0-1, and these values were inserted into another column in MS Excel. This function was conducted for as many articles there were that given day. By using a filter and listing the column consisting of RANDOM() from smallest to biggest numbers, the columns in the MS Excel were connected, and the computer provided a new value and changed the order of the articles in the column. Three articles from each day were selected by choosing the ones from the top.

![Image](image.png)

**Picture 1.** An example of conducting the randomization by using MS Excel.

Even though the selection of the articles was random, however, what the linguistic analyses are trying to identify in them are any linguistic features that deviate from Standard English, in terms of vocabulary, structure, style, grammar and syntax. By Standard English (SE) is meant the English that with respect to spelling, grammar, pronunciation, and vocabulary is substantially uniform though not devoid of regional differences, that is well established by usage in the formal and informal speech and writing of the educated, and that is widely recognized as acceptable wherever English is spoken and understood, (Merriam-Webster Online Dictionary, n.d.).

The study comprises micro-level linguistic analyses. It is important to understand the language of the media, as it comprises a great amount of the language people read and hear every day, therefore affecting and shaping people’s language usage as well. What is more, as social institutions, media also have a crucial role in shaping and reflecting, how for example culture and social life are formed and expressed (Bell, 1995). As celebrity journalism comprises a large proportion in the contemporary media world (Dubied & Hanitzsch, 2014), it is important to recognize how celebrity is constructed linguistically, as it affects how journalists use language and linguistic devices in their work. Furthermore, understanding its language usage is essential, as it can also help to understand the phenomenon of celebrity
itself. As both the celebrity news publications that are used in the study are solely online based, the analysis concentrates on the linguistic aspects of specifically online celebrity news coverage in the U.S. The analyses are conducted in the following manner:

1) Firstly, the articles are analyzed from a qualitative point of view, that is, the texts are categorized according to which kind of linguistic phenomenon they represent.
2) Secondly, the study applies critical linguistics and rhetorical analysis to examine the articles in terms of figures of speech, irony and style. In addition, the level of formality and informality of the language is evaluated as well, whether the articles contain, for example, slang words and face-to-face spoken conversation features.
3) Thirdly, the articles are investigated by applying corpus linguistics, which includes both quantitative and qualitative research approaches.

Both celebrity news websites are analyzed separately, as well as compared to each other. The methods are discussed in more detail in the following sections.

4.2 Linguistic Phenomena

Linguistics, in its most simple definition, is the scientific study of language. It can be structured in several different ways, depending on what kinds of aspects of language are studied. 1) Phonology is the study of sounds that deals with utterances in speech, 2) morphology involves the study of internal structure of words and phrases, 3) syntax explores the rules of a spoken sentence structure (e.g. word order) comparing it to the proper grammar that is the structural rules of language composition. 5) Semantics is the systematic study of meaning behind words and word combinations, and how language organizes and expresses meanings (Linguistic Society of America, n.d.). The study aims to identify and then categorize the language usage in the celebrity news articles according to these linguistic structures and aspects.

4.3 Critical Linguistics

Critical linguistics is a type of discourse analysis and it assists critical readers to identify patterns within language. Media discourse is one of the main concerns of critical linguistics, as it seeks to draw attention to the ways in which language is used in the media (Fairclough, 1995; Schröder, 2002).

Media discourse, on the other hand, refers to interactions that occur through a broadcast platform, either written or spoken, in which the discourse is orientated to a non-
present recipient (reader, listener or viewer). Media discourse is a public, manufactured on-recording form of interaction (O’Keeffe, 2012). The discourse of the media encapsulates three main aspects: the story (or spoken, written or visual text), the process involving the production of the text, and alignments to audience (Cotter, 2015). The first dimension of the discourse of the media, that is, the story or the text, is the main focus of this study. The texts of the celebrity news articles are considered from the point of view of discourse structure or linguistic function (Cotter, 2015).

The approach of critical linguistics addresses a close relationship between the linguistic details of media texts and the production of ideology. By studying the language patterns, the relationship between dominant and subordinate groupings can be charted, as well as the strategies which make such representations acceptable to the general public as “normal” (Faicloough, 1995; Schröder, 2002, Conboy, 2007).

The analyses of critical linguistics concern, for example, representation and the ideational function, how events, people and objects that are involved in them are represented grammatically (simple sentences). The basic principle is that coding events in language involves choices among the models, which the grammar makes available, and that such choices are potentially significant ideologically. That is, the choice of words or unspecified way of formulation may indicate that the words have certain ideological meaning (Faicloough, 1995).

As critical linguistics emphasizes the role of vocabulary in processes of categorization (including metaphors), lexical processes of the celebrity news articles are evaluated. In addition, the concept of modality (pronouns, modal auxiliaries, speech acts, among others) is used in critical linguistics to cover textual features that express writers’ attitudes towards themselves, their interlocutors and their subject-matters (Faicloough, 1995; Schröder, 2002).

Besides the lexical processes and modality, another concept within the framework of critical linguistics that is applied in the investigation of the celebrity news articles is implicature, which comprises a variety of communicative features, and it refers to the ability of reading between the lines of what is really meant in the text (Schröder, 2002).

Certain limitations have to be discussed concerning critical linguistics, though. The focus is upon the text and its productive practices, but texts are interpreted by the researcher without reference to the interpretative practices of audiences. Additionally, the linguistic analysis concentrates mainly on clauses, paying little attention to higher-level organization properties of the whole text. However, the limitations of critical linguistics
mainly reflect the shifts of focus and developments of the theory since the times when critical linguistics was developed, which was in the 1970s (Fairclough, 1995).

4.4. Rhetorical Analysis

Rhetoric and Style

The original model of rhetoric, the art of persuasion, was developed in ancient Greece. It is a practical skill that is used to persuade the audience. Ancient Greeks invented a five-step process that would make the speaker deliver a believable and persuasive speech: *invention/idea, arrangement, style, memory and delivery* (Henriksen, 2011).

Aristotle then defined rhetoric as the “power of discovering the means of persuasion in any given situation” (as cited in Kuypers & King, 2009, p. 2). Since the beginning of the 21st century, the study and definitions of rhetoric have expanded significantly. Kuypers & King (2009) define rhetoric as “the strategic use of communication, oral or written, to achieve specifiable goals” (p. 4). This definition expresses two main ideas; the strategic (or intentional) nature of the language that is used, and knowing the goals that are wished to achieve through the used language (Kuypers & King, 2009). Rhetoric is thus closely linked to linguistics, as words, ways of speaking and shaping the message can have a fundamental impact on how the message is received and understood (Henriksen, 2011). Therefore, a rhetorical analysis, which is also conducted in this study, is a suitable complement to the study of linguistic features in terms of methodological compatibility, as rhetorical analysis seeks to discover language usage as well.

Rhetoric is strategic since it is intentional. Communicators (who aim to control the way their message is received) plan beforehand what to say and how to formulate it. When rhetoric is used this way, the communicator also provides reasons why the receiver should agree with them. In addition, rhetoric has informative and persuasive element. It aims to affect the personal and collective behavior of the receiver by having him or her voluntarily agreeing with what is said. Rhetoric works towards a goal. It may involve trying to make the receiver believe in a certain way, or it can work toward the enacting of one course of action over another. The trick of a person trying to persuade someone is to assure him or her that the level of probability is high enough to convince the receiver that a certain matter or course of action will “probably” be better than any other course of action (Kuypers & King, 2009).

The rhetorical devices that this study focuses on when analyzing the celebrity news coverage are those of style, or *elocutio* in rhetorical terms. Style in rhetoric is concerned with how to form an argument, that is, how to make the text as persuasive as possible. Style is
central in communication and persuasion, since the way the author or a speaker shapes and styles a message can have a major influence on how the receiver comprehends it, as stated before (Henriksen, 2011). The rhetorical analysis that is conducted in this study has an analytical approach, as it focuses on the discourses of the celebrity news articles. The analysis concentrates on, how figures of speech (figurative language) are used in the celebrity news articles with a certain purpose. The stylistic aspects that are investigated in this study are pronouns, and irony as a figure of speech.

**Pronouns**

The meaning of a text is produced in a complex negotiation between the author and the reader. When the author of a text addresses to the reader by the use of “you”, or refers to themselves by using “we”, it is called direct addressing. This stylistic feature is used often in commercials and advertisements (Montgomery et al, 2013). For instance, a phrase that TMZ uses often in its articles “we broke the story”, or a question that Radar Online poses frequently “what do you think of…” are examples of directly addressing to the reader.

The use of the pronoun “we” can function in two different ways, inclusively and exclusively. In terms of the celebrity news articles, “we” may refer to the author of the article and other staff member of the celebrity news outlet (exclusive “we”), or it may refer to the people of the celebrity news outlet and their readers (inclusive “we”). By using the pronoun “you”, on the other hand, the author generally separates himself or herself from the reader (Henriksen, 2011).

*Spoken narrative,* on the other hand, specializes in strategies for capturing speech. The author can for example mimic tones of voice and pitch in order to deliver emotions, reactions and even to distinguish one speaker from another (Montgomery et al, 2013).

**Irony**

Irony is a use of language in which what is said or written is not intended to be taken seriously or as sincere. It includes the figurative or indirect interpretation of an utterance, text, event or situation. It includes something that the interpreter already knows or has been led to believe, which encourages him or her not to treat the subject at face value. Irony has various forms, of which the most distinguished type in texts is called verbal irony, that is, irony that is associated with verbal communication (Montgomery et al, 2013).
A communicated meaning can be analyzed into two different components: a proposition and an attitude towards the proposition. A proposition is a sentence that makes a statement of any kind, whilst the attitude that one has towards the proposition is called propositional attitude. The most common attitude one has towards a proposition is belief. In other words, one tends to believe what is said (Montgomery et al, 2013).

However, one can also express propositions that one does not believe, in which case the propositional attitude is concealed and the person is lying. If the communicator signals in some way that he or she does not believe the statement to be true, the communicator is being ironic (the communicator communicates a proposition while simultaneously signaling that he or she does not believe the proposition to be true). In effect, in ironic situations, the opposite of what is actually said, is implied. In the proposition itself there is no difference: the attitude that is communicated along with the proposition is the most important feature (Montgomery et al, 2013).

Verbal irony is most successful when the communicator provides enough evidence to show that his or her attitude is that of disbelief, rather than the expected attitude of belief. Disbelief towards what is being said is not the normal state; consequently, there must be something unusual or strange about the text to indicate that the communicator disbelieves what he or she is saying. For example, signals of verbal irony are obvious exaggeration and overemphasis of a proposition, and a clear contradiction between what is said and what is already known. All varieties of verbal irony function in the same way, as they are indicating that the proposition is not believed by the communicator, and they should not be believed by the receiver either (Montgomery et al, 2013).

In sarcasm, on the other hand, which is often called the lowest form of verbal irony, an exaggerated tone communicates the attitude of disbelief. However, in sarcasm the disbelieved attitude remains indirect: the communicator does not signal directly that he or she does not believe what he or she is saying (Montgomery et al, 2013).

Ironic usually involves certain tension between what is actually said and meant, thus it opens up a subtle “game” between the communicator and receiver (Montgomery et al, 2013).

What is more, a writer ordinarily has an “attitude” to his or her reader. The tone of his or her utterance reflects his or her awareness of this relation and the sense of how he or she stands towards those that are being addressed. A more expansive definition of tone can also be broadened to include writer’s attitude towards his or her subject, as well as the addressees or readers. The writer can adopt a playful tone towards a subject, demonstrating
that he or she does not take the subject too seriously, a *serious tone* showing that the writer believes that the subject merits careful reflection, a *neutral/dispassionate tone* demonstrating a lack of emotional involvement, or a *sarcastic tone*, which indicates a negative attitude towards the subject (Jasinski, 2001).

### 4.5 Corpus Linguistics

“*Corpus linguistics* is the study of language based on examples of real life language use” (Baker, 2006, p. 1). It combines approaches from both qualitative and quantitative techniques: unlike purely qualitative research approaches, corpus linguistics utilizes bodies of electronically coded texts, therefore, it implements a more quantitative methodology, for instance by using frequency information about how often certain linguistic phenomena occur in a text. Investigating association patterns in a text (that is, measuring the extent to which variants and features are associated with contextual factors) represent a quantitative approach, whereas functional interpretation of the text is an important part of a qualitative corpus-based analysis (Baker, 2006).

Corpora are usually large, consisting of thousands or even millions of words. They represent samples of a specific type of language that occurs naturally, in real life situations, thus they can be used as standard references of how language is used. The language data of corpora is stored on computers. As this study analyzes only 180 articles, the total word count of the corpora is rather small. However, corpus linguistics can make the identification of patterns and word frequencies and the like considerably easier than conducting it manually; since corpora can carry large amount of texts, it can reveal linguistic patterns and information about frequency, among other things, that otherwise would take days or weeks to uncover by hand (Baker, 2006).

As corpus linguistics involves both quantitative and qualitative approaches, it enables the analysis of discourses (systems of statements constructing an object). The corpus processes (computer procedures that manipulate the data in various ways) uncover linguistic patterns, making it possible to comprehend how language is used in the construction of discourses (Baker, 2006). One way that discourses are constructed is through language. Even though language (as an abstract system: phonetics, grammar, lexicon etc. and as a context-based system of communication) is not the same as discourse, but analyses of language in texts can be carried out in order to reveal traces of discourses (Baker, 2006).

Baker (2006) argues that corpus linguistics provides different ways to analyze the data. The linguistic patterns can be investigated by using several techniques: keywords,
collocations, frequency lists, clusters, and so forth. These concepts are introduced in more
detail in the following section. For every used technique limitations need to be set, for
example, how many times a certain word has to occur in the corpus in order to decide that the
word is significant enough and worth of investigation. External criteria, such as the length of
the examined texts could be used to set such boundaries.

4.5.1 Compiling the Corpora
For the electronically coded texts, software called AntConc is used in this study. AntConc is
“a freeware, multiplatform tool for carrying out corpus linguistics research and data-driven
learning” (Anthony, 2014). It is created by Laurence Anthony. It is used by inputting the
wanted search term(s) in the software and choosing from the variety of tools, what kinds of
aspects of the texts are investigated (Anthony, 2014).

For the analyses of the celebrity news outlets, two specialized corpora were
compiled. A specialized corpus is used to study aspects of a certain variety or genre of
language, for example the language of spoken communication or the language of news
articles. Consequently, for a specialized corpus only texts that conform and fit to this
specialized area are collected. A specialized corpus can also be limited to certain topic and the
texts are chosen merely according to the topic in question (Baker, 2006).

In the case of this study, the specialized topic comprehends celebrity news. Since the study investigates two celebrity news websites, two specialized corpora were
compiled: a corpus that consists of 90 articles from Radar Online and a corpus that consists of
90 articles from TMZ. The chosen articles that were taken from the websites were converted
into .txt-files by using AntFileConverter, as AntConc cannot read PDF and Word files. Once
the articles were converted into a proper form, they were simply opened in the AntConc
software (Anthony, 2014). By using AntConc, the two corpora can be investigated
simultaneously in order to see what kinds of linguistic phenomena and features occur in the
texts as a whole. In addition, a corpus consisting of Radar Online articles and a corpus
consisting of TMZ articles can also be analyzed separately, which allows to make
comparisons between the two corpora and the language that are used in the websites.

By using Antconc, the following tools and techniques are used to investigate the celebrity
news articles:

• Concordance: A concordance is a list of all the occurrences of a particular
search term in a corpus and it is presented in the context in which they occur,
usually a few words to the left and right of the search term in question (Baker, 2006).

- Collocation: When certain words regularly appear near each other and their relationship is somehow statistically significant, these co-occurrences are called collocates, and the phenomenon of such words frequently occurring next to or near other is called collocation. Collocation is a way of understanding meanings and associations between words that are difficult to establish otherwise from a small-scale analysis of a single text (Baker, 2006).

- Word list: This tool in Antconc counts all the words in the corpus and shows them in an ordered list, making it possible to locate quickly which are the words that are the most frequent in the corpus (Anthony, 2014).

- Clusters: Using clusters of words, patterns and groups of words can be searched. It shows words that result together with the wanted search term (Anthony, 2014).

- Keyword list: This tool shows which words are unusually frequent or infrequent in a corpus, when compared to words of a reference corpus (Anthony, 2014).

*Reference corpus* is another type of corpus that is often very useful for corpus-based analyses of discourses, although it usually does not play the most important role in the research. Reference corpus consists of a large corpus (it can contain millions of words from a wide variety of texts), which is representative and demonstrative of a particular language variety. For example, the British National Corpus (BNC) is a reference corpus that consists of approximately one hundred million words of spoken and written data, which has been gathered from newspapers, journals, popular fictions and academic books, informal conversations, and so forth. The benefit of a reference corpus in corpus linguistics is that since the reference corpora are generally large and illustrative enough of a particular language or genre of language, they can be used themselves to find certain discourses. In addition, reference corpora function as a good benchmark of what is typical in language, by which the data from the researcher’s own corpus is compared to (Baker, 2006).

In this study, as stated above, two small specialized corpora are built (*Radar Online* and *TMZ*). What is more, a reference corpus is used to identify hypotheses about the language that are formed during the investigation of the specialized corpora, and these findings are compared between the reference corpus and specialized corpora. The reference
corpus that is used in this study is the Corpus of Contemporary American English (COCA), one of the most widely-used corpora of English. It contains more than five hundred million words and its data is equally divided between spoken language, fiction, popular magazines, academic texts, journals and newspapers. The data has been gathered in 1990-2017. The corpus is accessible online for free (Corpus of Contemporary American English, n.d.).

COCA was selected as the reference corpus since the corpus contains, as the name of the corpus suggests, American English, and the celebrity news outlets that are the focus of this study are also American based. COCA was selected for the study despite the fact that the Radar Online and TMZ articles are from 2018. However, as the newest material in COCA is from 2017, it is seen to be current enough to compare with the contemporary language usage.

Advantages and Limitations of Corpus Linguistics
Corpus-based analysis provides several advantages to analyze discourses. For example, corpus linguistics can help to restrict bias, as it is harder to be selective about the research material when there can be dozens or hundreds of texts to examine electronically. In addition, with a corpus analysis, the researcher usually starts from a position where the data has not been chosen to confirm conscious or subconscious hypotheses (Baker, 2006). Baker (2006) argues that in corpus-based analysis, there are usually exceptions to any patterns or rule. He states that it is important to report these findings alongside the other results, even though the exceptions do not need to be over-reported.

Besides helping to restrict bias, corpus linguistics has other advantageous ways to approach discourses. One of the most central ways that discourses are circulated and strengthened in society is through language usage, and the mission of a discourse analyst is to uncover, how language is employed, in order to reveal underlying discourses. By being more aware of how language is constructed and drawn on to build discourses or ways of looking at the world, it is easier to resist attempts by writers of any texts, how they might try to persuade or manipulate people by suggesting what is common sense or accepted truth. However, it can be difficult to tell, whether a discourse is typical or not. This is where corpus linguistics and corpus-based analysis are useful. Associations between words, occurring repetitively in natural language use, provide better evidence for underlying hegemonic discourses than looking at a single case (Baker, 2006).

According to Baker (2006), while corpus linguistics can act as an extremely useful method of analyzing discourses, certain limitations have to be taken into account.
Firstly, as the corpus data does not interpret itself, it is the researcher’s task to make sense of the language patterns that are found in the corpus, postulating reasons for their existence, or looking for more evidence to support hypotheses. The researcher’s findings are always interpretations, and that is why restricting bias can only be restricted, not removed completely.

Additionally, a corpus-based analysis tends to put focus on patterns and frequencies, although frequent language patterns do not always necessarily imply underlying discourses. Also, frequent language patterns do not always imply mainstream ways of thinking: sometimes what is not said is more important than what has been said. Furthermore, corpus linguistics has to take into account that the meanings of words can change over time, and certain words have different meanings and triggers for different people (Baker, 2006).

4.6 Reliability and Validity

The concepts of reliability and validity are briefly presented in qualitative and quantitative researches, as the methods in this study apply both qualitative and quantitative research approaches.

In quantitative research, where reliability and validity have largely been developed, reliability is concerned with whether the results of the study are repeatable. Reliability can be further defined by the concepts of stability, internal reliability and inter-observer consistency. Stability checks whether a measure is stable over time, so that the results do not alternate, and internal reliability is concerned with whether the indicators are consistent. Inter-observer consistency deals with situations when subjective judgment is involved in categorization of data, and more than one observer is involved in processing the data, when the decisions might not be consistent. Replicability is the degree to which the results of the study can be reproduced. Validity, on the other hand, is concerned with the integrity of the conclusions of the study. It means whether an indicator that is planned to evaluate a concept really measures it (Bryman, 2012).

In qualitative study, reliability and validity have a very close meaning than in quantitative research. External reliability refers to the degree to which a study can be replicated and internal reliability is concerned with a case where there is more than one observer, whether they agree to their observations or not. Validity, then, refers to whether the researcher is observing, measuring and identifying what they say they are (Bryman, 2012).

The results of this study are seen reliable as the study has a relatively objective listing of linguistic categories, and the results are particularly replicable in the domain of
corpus linguistics. However, some aspects of the study may have concerns. For example, there is always room for subjectivity when it comes to deciding to which category each linguistic phenomenon found in the data belongs to. This was especially the case of classifying irony and other rhetorical devices. However, this was seen as a lesser concern, as using multiple research methods minimized the issue of subjectivity.

Triangulation (using multiple methods of analysis) was employed in the study. Triangulation has several benefits, for example it facilitates validity checks of hypotheses, it allows the researcher to respond more flexibly to unexpected problems and it anchors results in more robust explanations and interpretations (Baker, 2006).

All the methods of the study (critical linguistics, rhetorical analysis and corpus linguistics, as well as identification of linguistic phenomena) are applied to complement each other in terms of cross-validating the findings, and the similarity of the findings retrieved with the help of different methods testifies to the overall validity of the study design.

5 Results and Analysis
5.1 Identifying the Different Types of Linguistic Phenomena
The first analysis of the celebrity news articles concerned the identification of the linguistic phenomena that are used in them. In the qualitative analysis the following linguistic aspects in the celebrity news articles are investigated:

1) Slang, unconventional spelling, abbreviations, taboo language and colloquial expressions (linguistic aspect of lexicon/vocabulary)
2) Unconventional grammar and sentence structure (linguistic aspect of syntax and grammar).
3) Onomatopoeia (linguistic aspect of phonology).

The aspects above form the structure of linguistics, as mentioned previously; therefore, they form the basis for the analysis of categorizing the linguistic phenomena. However, the aspect of semantics is not included in this analysis. Each of the linguistic aspect is presented in the following paragraphs, and for each linguistic category certain choices in terms of classification have been made, as arriving at this categorization was not easy, since some of the phenomena found in the data could have been classified into more than one category. This is explained and discussed in the following sections. The framework of critical linguistics has
also been applied in the categorization of the linguistic phenomena of the celebrity news articles.

5.2 The Characteristics of the Celebrity News Articles

1) Lexicon and Vocabulary

The first identified category in the data concerns lexicon and special vocabulary. It is further divided into the following subcategories:

- **Unconventional spelling forms.** This subcategory represents word forms that are spelt in a non-conventional way, offering variation to standard spelling. **Word blending** (e.g. wanna instead of want to), **g-dropping** (e.g. shakin’ instead of shaking) and **abbreviations** are cases that are classified as unconventional spelling forms. Abbreviations are shortened forms of a word or phrase (Baron, 2003). Acronym is a type of abbreviation that is formed from the initial letters of other words and it is pronounced as a word (Oxford Dictionaries, n.d.), for example YOLO that stands for you only live once. Another type of abbreviation is an initialism, which consists of initial letters of words that are pronounced separately (Oxford Dictionaries, n.d.), for example btw, the initials of by the way.

- **Colloquial expressions and slang.** Colloquialisms, colloquial expressions and slang are words or phrases that are not formal or literary forms of Standard English, but are often used in familiar or ordinary conversation, for example a cop instead of a police officer (Oxford Dictionaries, n.d.).

- **Taboo language.** Taboo language is classified as swearing, sexual connotations and references to body parts and bodily excretions (Gao, 2013). Swearing can be further divided into annoyance swearing, where the swearer is stressed, and social swearing, where the swearer is relaxed, and it is associated with social context (Kapoor, 2016).

It has to be noted though, that some of the classifications of the vocabulary could have been categorized into another category as well. In particular, certain cases of the unconventional spelling forms, colloquialisms and slang words could be seen as interchangeable. For example, the word *bro* which stands for *brother*, can be identified as slang and as an unconventional spelling form.

The meanings of all the colloquial and slang words and phrases in the examples have been checked by using the Online Slang Dictionary of American English and Urban slang (The Online Slang Dictionary).
<table>
<thead>
<tr>
<th>Celebrity Outlet</th>
<th>Vocabulary/Lexicon</th>
<th>Unconventional Spelling Forms</th>
<th>Colloquialisms &amp; Slang</th>
<th>Taboo Language</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Radar Online</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>hubby</em> (husband)</td>
<td></td>
<td><em>pal</em></td>
<td><em>So f**king what?</em></td>
</tr>
<tr>
<td></td>
<td><em>Calif</em> (California)</td>
<td></td>
<td><em>cops</em></td>
<td><em>bull f**king sh</em>t*</td>
</tr>
<tr>
<td></td>
<td><em>kinda</em> (kind of)</td>
<td></td>
<td><em>scumbag</em></td>
<td><em>She has great tits and I know her p</em>**y’s bomb too.*</td>
</tr>
<tr>
<td></td>
<td><em>BF</em> (boyfriend)</td>
<td></td>
<td><em>gung-ho</em> (enthusiastic)</td>
<td><em>And everyone else can f**k off.</em></td>
</tr>
<tr>
<td></td>
<td><em>u/ur</em> (you/your)</td>
<td></td>
<td><em>hoe</em> (whore)</td>
<td><em>I don’t want people to sh</em>t talk you*</td>
</tr>
<tr>
<td></td>
<td><em>lib</em> (liberation, liberal)</td>
<td></td>
<td><em>dud</em> (a failure)</td>
<td><em>Get out of my f</em>**ing house*</td>
</tr>
<tr>
<td><strong>TMZ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>the Prez</em> (the President)</td>
<td></td>
<td><em>bitch</em></td>
<td><em>That's my motherf</em>**ing daughter*</td>
</tr>
<tr>
<td></td>
<td><em>do</em> (hairdo)</td>
<td></td>
<td><em>swag</em></td>
<td><em>Hey, I’m burning your s**t</em></td>
</tr>
<tr>
<td></td>
<td><em>FYI</em> (for your information)</td>
<td></td>
<td><em>shindig</em> (a small party)</td>
<td><em>aka 30 seconds to crap your pants!</em></td>
</tr>
<tr>
<td></td>
<td>‘<em>splainin</em>’ (explaining)</td>
<td></td>
<td><em>dude</em></td>
<td>‘<em>cause why the hell not?</em></td>
</tr>
<tr>
<td></td>
<td><em>outta</em> (out of)</td>
<td></td>
<td><em>whacked</em> (strike)</td>
<td><em>The dynamic boob duo</em></td>
</tr>
<tr>
<td></td>
<td><em>ya</em> (you)</td>
<td></td>
<td><em>sucking face</em> (kissing)</td>
<td><em>Iggy dumped a bunch of Nick’s crap</em></td>
</tr>
</tbody>
</table>

As it can be seen from table 1, all the subcategories presented above are present in both celebrity news outlets. In the column of taboo language, the swear words are censored as this is the way they were spelled in the articles. The unconventional spelling forms *u/ur* in the *Radar Online* column are **letter homophones**. Letter homophones are abbreviations that consist of a single letter that are replacing the whole word, such as *u* instead of *you* (for example, *how ru*, meaning *how are you*). The purpose of the letter homophones is to create approximately the same phonetic rendering and phonetic spelling of the word (Merchant, 2001). Because of the phonetic spelling style of the word, letter homophones could be classified in the category of phonology as well, but this study has chosen to analyze letter homophones strictly from the perspective of spelling, rather than pronunciation. The feature of letter homophones could not been found in the *TMZ* articles. The choices of the particular vocabulary (lexical processes) indicate that specific kinds of words are typical for celebrity
journalism. All the subcategories of vocabulary that were presented above are connected to “everyday life” (Conboy, 2014) and they are unsophisticated and simplistic. Examples (1), (2) and (3) demonstrate more of the usage of the special vocabulary in the articles.

(1): “This doesn’t seem like a baptism to me. This just seems like a hoe bath.” (Radar Online, colloquialism & slang)

(2): ...and he's got a helluva explanation for why. (TMZ, word blending of hell of a, which could also be categorized as taboo language)

(3): …the dude sucks at hoops and has a snowball's chance in hell of making the NBA. (TMZ, colloquialism & slang used together with a swear word)

2) Grammar and Syntax
The second found category of the celebrity news articles involves words, phrases and clauses that do not follow the standard structural forms of language composition.

Table 2. Examples of unconventional grammar and syntax in the celebrity news articles.

<table>
<thead>
<tr>
<th>Celebrity Outlet</th>
<th>Unconventional Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radar Online</td>
<td>-</td>
</tr>
<tr>
<td>TMZ</td>
<td>who woulda thunk</td>
</tr>
<tr>
<td></td>
<td>she knows who dunnit</td>
</tr>
<tr>
<td></td>
<td>we ain't talking about the Jets</td>
</tr>
<tr>
<td></td>
<td>You sure you want that, Anthony Joshua?</td>
</tr>
<tr>
<td></td>
<td>the guys made sure to throw it back in the water after bro’in out on their find.</td>
</tr>
<tr>
<td></td>
<td>We're told Jamie Spears ain't budging ...</td>
</tr>
</tbody>
</table>

Table 2 demonstrates that the unconventional grammar and syntax formations are typical only in the TMZ articles. Examples (4), (5) and (6) illustrate their usage in more detail.

(4): ...who woulda thunk a simple night out on the town with her hot soccer boyfriend would turn into an awesome, belated surprise birthday party (TMZ, word blending of would have and an unconventional past tense formation of the verb thought)
(5): Chrissy's been dropping more than a few hints she knows who *dunnit* (TMZ, word blending of *done it* and absence of the auxiliary verb *have*)

(6): …but not Sly's wife, Jennifer Flavin, who says *it's really ain't no thang!* (TMZ, unconventional sentence structure, double negation, non-standard form of *is not* and a slang form of *thing*)

3) Phonology

In this third category are included the cases of *onomatopoeia*, that is, exclamatory spellings of words that imitate a production of a sound (Thurairaj et al, 2015). Other included subcategories are *sound representations* and *shouting*. Sound representation in this study comprehends the lengthening the spelling of a word in a playful manner (Herring, Stein & Virtanen, 2013), for example the vowels in the word “*sooooooo*” to resemble its exaggerated pronunciation in speech, and other similar spoken language elements. As the cases of onomatopoeia and sound representations are very similar, they are presented under the same column in table 3. Shouting is considered as capitalized writing.

**Table 3.** Examples of the usage of phonology in the celebrity news articles.

<table>
<thead>
<tr>
<th>Celebrity Outlet</th>
<th>Phonology</th>
<th>Onomatopoeia &amp; Sound Representations</th>
<th>Shouting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radar Online</td>
<td>aw</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>errrrr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>woohoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uh Oh!</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ok sooooo</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>READ THE SHOCKING COURT DOCUMENTS HERE!</td>
<td></td>
</tr>
<tr>
<td>TMZ</td>
<td>BDRRRAAPP!</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>yikes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ummm, yeah</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>wepa!</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>oh</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIIIAAAAN</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>THIS KID IS AMAZING!!!!</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GOING PRO, BABY!!!</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>YAAASS!</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>it's about to get ALL SURGICAL UP IN HERE!!!</td>
<td></td>
</tr>
</tbody>
</table>
As table 3 shows, all subcategories can be found in both celebrity news outlets, however, shouting is a lot more rare in *Radar Online*, as only one case was located. Examples of the usage of phonology are provided below.

(7): *Nothing says "Wepa!" like the open seas.* (TMZ, onomatopoeia)

(8): *Looks like Desiigner will skate on this one.* "BDRRRAAPP!" (TMZ, onomatopoeia)

(9): *Woohoo! It’s here.* (Radar Online, onomatopoeia)

5.3 Critical Linguistics and Rhetorical Style

The next part of the analysis of the celebrity news articles concerned critical linguistics and the rhetorical devices of style. The following examples demonstrate how the writers of the celebrity news articles employ rhetorical devices in terms of pronouns and irony. Both cases of direct addressing (you vs. we) can be located in both celebrity news outlets. In the article examples below the tone of the author is evaluated as well, as the authors add frequently their comments and opinions of the subjects. This applies to both *Radar Online* and *TMZ*. Another aspect that is examined is the usage of verbal irony. Additionally, the concept of modality (for example pronouns, auxiliary verbs and speech acts), which is closely related to rhetoric, is applied within the framework of critical linguistics to uncover features that express the authors’ attitudes towards themselves and their subjects.

**Examples of Direct Address: RADAR ONLINE**

(10) *What do you think about Joe’s racy sex talk? Sound off in the comments below!*

(11) *RadarOnline.com has learned exclusively!*

(12) *Stay with Radar for updates to the story.*

(13) *As Radar previously reported, Moore is on the chopping block for her secrecy and diva behavior.*

(14) *As Radar readers know, Wilkinson, 32, and Baskett, 35, have been having issues for quite some time.*

As it can be seen from the examples above, *Radar Online* addresses in a direct manner to their readers by the pronoun “you”, and occasionally asks the readers to leave comments as well, as if they were having a dialogue. When addressing to the writers themselves (and the website), instead of using the pronoun “we”, they refer to themselves by using the name of the website *Radar* or *RadarOnline.com*. In addition, they use presuppositions by assuming
that the writer and readers already share certain knowledge, for instance “As Radar readers know”, which creates a feeling of unity.

**Examples of Direct Address: TMZ**

(15) *As we reported*, the Prez [President Trump] checked out some wall, and at a news conference he called Mexicans "incredible climbers." Classic 45, **right?**

(16) Bella Thorne and her musician boyfriend went ring shopping together ... but it's **not what you think.**

(17) And just for old times' sake ... here it is again. **Try to resist jamming out, if you dare.**

(18) *As we first reported* ... Angie filed divorce docs last month after about 4 years of marriage.

(19) Her choice of words is priceless, and has a lot to do with glass toys. **Just watch (headphones would be best), and remember** ... Mary ran, seriously, for governor of California.

As the examples show, TMZ writers address to themselves by the pronoun “we”, unlike Radar Online writers. In addition, TMZ writers address directly to the reader by the pronoun “you”, but they also have direct advice for the reader, as if having a conversation with them, for instance examples (17) and (19). This also creates a feeling of unity between the writers and readers.

**Examples of Tone: RADAR ONLINE**

(20) **Too bad** Kris doesn’t think the coincidence is a laughing matter.

(21) **The only problem?** Kunis nixed the idea!

(22) **Busted!**

(23) **But he won’t miss it again!**

(24) **Time for “the talk!”**

The modality of these examples indicates that the writers of the articles include their own opinions and comments in them and make humoristic comments, bringing conversational elements to the writing. The tones of these examples can be seen as playful, showing that the authors do not take the subjects too seriously. The writing style is thus very subjective. These fragments are also examples of very simple sentences that are typical for this writing style.
Examples of Tone: TMZ

(25) Of course, the museum steps and statue are part of the Rocky legend -- it's where Balboa proved to himself that he had the heart, the will and the EYE OF THE TIGER in multiple 'Rocky' movies.

(26) Oddly specific ... but hey, he got it done. Bienvenido al baile, Mr. Smith.

(27) How dope [good] is medicine?!

(28) No violence that time either, but man, does Breezy have parking lot issues!

(29) Interestingly, Carl opted to file his response to Angie's divorce petition all by himself -- look, Ma ... no lawyers! We're guessing he can more than afford one, though.

As it can be noted from the modality of the examples above, TMZ writers add their own comments and opinions frequently as well, making the writing very subjective and colorful. Example (29) shows how the writer not only adds his or her own comments, but also switches between the narration of the events and the speech of another person that they are reporting on. This is an example of spoken narration. The tones of the authors in these examples are playful, as they are humoristic. Example (27) is also a case of a rhetorical question. Rhetorical question can be defined as a posed query for which a response is not expected. The form of the question provides for a sense of openness because multiple answers are possible, but in most cases the situated force of the utterance undermines the veracity of any answer other than the one that is implied by the writer (Jasinski, 2001).

Examples of Verbal Irony: RADAR ONLINE

(30) What? Dated?

Examples of Verbal Irony: TMZ

(31) He might not brag about this one, but Lindsay Lohan famously wore Givenchy back in 2012 ... during one of her many, many, many court appearances.

(32) And they lived happily ever after...for now.

(33) President Trump would like to introduce you to the 2018 White House spring interns, who -- when it comes to diversity -- pale in comparison to the last President's. #WhiteHouseSoWhite
Interesting note -- Tiffany Haddish just became a spokesperson for Groupon in January. She's yet to comment on the n-word controversy, but ya gotta think she'll have something a lot to say about it.

In fact, Jenn says what all us normies gotta do to be like Sly is start hitting the gym! Right ...

The examples above designate that the irony in the celebrity news reporting is more common in the TMZ articles, as only one clear example of Radar Online was located. The extracts from TMZ show cases that include certain tension between what is actually said and meant, opening up a subtle “game” between the author and reader. For example, the author is overemphasizing and exaggerating the proposition in example (31), and contradicting what is actually said and meant in example (35). What is more, the tone in example (32) is not very sincere.

The only illustration from Radar Online, example (30) shows a case of disbelief of the author, and disbelief towards what is being said is not the normal state, therefore, there must be something unusual or strange about the statement. Example (30) is also another case of a rhetorical question. In example (34), on the other hand, the writer is having a little word play with the words something and a lot, and thus turning it into an ironic joke.

It can be stated that the tone of most of the article examples presented above is playful; however, in example (33) the tone of the author can be seen as sarcastic; there is an indication of a negative attitude of the subject.

In terms of implicature (reading between the lines what is really meant), even though some of the examples above show irony and contradicting propositions, the language is simple-minded enough to understand what the author means. In other words, the reader is not required to interpret the real meaning of a proposition, as the writing and word choices are formulated so that there are no doubts about their meaning. The irony and contradicting propositions in the article examples are meant in a humoristic way.

Furthermore, the representations of most of the examples above illustrate very simple sentences. The word choices represent vernacularism that is typical for this news genre (Conboy, 2014)
5.4 Corpus-Based Analysis

1) Frequency, Word List and Keyword List

Even though frequency is one of the most basic and central concepts of corpus analysis, it should not play the most important part of the analysis, as frequencies can be generalizing, reductive and distract from more interesting data interpretations. When frequencies are used sensitively though, they can uncover a variety of interesting phenomena. Frequencies are important because language is not random. Languages are based on rules, they consist of patterns and certain words occur in relationship to other words. However, people normally have some kind of choice about their language usage. The relation of these two affairs, language as a set of rules and language as a free choice, make the concept of frequency so important: when people speak or write in an unexpected way, or make a linguistic choice over another, then that reveals something about their intentions, whether it is conscious or not (Baker, 2006).

As mentioned previously, the amount of articles is 90 per celebrity news website, thus 180 in total. After compiling the specialized corpora, it can be seen that the total amount of tokens (words) in the Radar Online corpus is 31448. From here on, the initialism RO is used when referring to the corpus. The total amount of tokens (words) in the TMZ corpus 15180.

Firstly, as it can be seen from the word tokens, these are very small corpora. Secondly, as the numbers indicate, the amount of words is remarkably higher in RO, almost twice the amount of TMZ words. This difference can be explained by the fact that even though the amount of the articles of the websites is the same, the TMZ articles tend to be very short, they often consist of only a few paragraphs, whereas the Radar Online articles are much longer; they often consist of at least one to two A4 pages. Thirdly, when using the Word List –tool in AntConc for calculating the most frequently used words in the corpora, it is shown that the most frequently used words are mainly grammatical words, such as the, to, and, etc. These results are summarized in the table below.
Table 4. The total amount of word tokens of the corpora and the most frequently used words of the corpora by using the Word List-tool.

<table>
<thead>
<tr>
<th>Specialized Corpus</th>
<th>Total amount of words</th>
<th>The most frequently used words</th>
</tr>
</thead>
<tbody>
<tr>
<td>RO</td>
<td>31448</td>
<td>the, to, and, a, in, of</td>
</tr>
<tr>
<td>TMZ</td>
<td>15180</td>
<td>the, to, and, a, in, he</td>
</tr>
</tbody>
</table>

It has to be taken into account that almost all forms of language contain a lot of grammatical words (Baker, 2006), which can also be seen in the results of these two specialized corpora. In order to evaluate if these grammatical words appear more often than expected, the corpora are compared to the reference corpus of Corpus of Contemporary American English (COCA). The comparison is conducted by calculating percentage frequencies.

When comparing these grammatical words to COCA, the following results are achieved. Even though the COCA corpus consists of millions of words, thus it is significantly larger than the RO and TMZ corpora, the percentages can still give reference value to the investigation.

Table 5. Percentage frequencies of some of the most frequently occurring grammatical words in the specialized corpora and their equivalencies in the reference corpus.

<table>
<thead>
<tr>
<th>Word</th>
<th>% frequency RO</th>
<th>% frequency TMZ</th>
<th>% frequency COCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td>4,43</td>
<td>4,70</td>
<td>5,42</td>
</tr>
<tr>
<td>To</td>
<td>2,75</td>
<td>2,54</td>
<td>2,55</td>
</tr>
<tr>
<td>And</td>
<td>2,45</td>
<td>2,53</td>
<td>2,69</td>
</tr>
<tr>
<td>A</td>
<td>2,14</td>
<td>2,42</td>
<td>2,27</td>
</tr>
<tr>
<td>In</td>
<td>1,78</td>
<td>1,74</td>
<td>1,82</td>
</tr>
</tbody>
</table>

As it can be seen from the table, the percentages follow the same patterns. Even though the percentage frequencies vary between the corpora, the differences in the percentages are not significant. When the grammatical words of the RO and TMZ corpora are checked against the frequencies of the reference corpus, it can be noted that the grammatical words in the two specialized corpora do not occur more frequently than expected.

The Word List –tool shows that the top lexical collocates that carry meaning in the RO corpus are her, was, she, he, is, that is, pronouns and different forms of the verb be,
but even these words do not have that much meaning per se. The first actual noun on the word list is radar, referring to the website itself, which is mentioned 123 times and it is ranked as the 29th most frequently occurring word. The next noun and a word that carries meaning on the list is star, which is mentioned 90 times and ranked as the 42nd most used word.

When using the Word List – tool in the TMZ corpus, it can be seen that the top lexical words that carry meaning are similar: he, was, his and her, but mainly the most frequently used words are pronouns, auxiliary verbs, articles or prepositions. The first word that represents neither of the previous categories is trump, which occurs 35 times and is ranked as the 48th most frequent word. The next collocates that carry meaning are back, occurring 30 times, and tmz, which occurs 29 times and is ranked as 62nd most used word in the TMZ corpus.

The Keyword List –tool uncovers the words that are significantly more frequent or infrequent in both corpora when compared against each other. As it can be seen from the pictures below, both corpora have their own name, radar/radaronline and tmz as one of the top words, which is understandable as neither of the websites is referring to each other, but referring to themselves. The pictures also show that in the RO corpus the word said is more frequent than in the TMZ corpus. However, when looking at the keyword list in the TMZ corpus, it can be noticed that TMZ uses the same verb, but in a present tense: says.

When examining the keyword lists (with the focus on words that carry meaning), it can be seen that the words exclusively, revealed, source and documents appear in the RO corpus more frequently than in the TMZ corpus. In addition, two names, Shannon and Jenner are on this list. Based on these top keywords, and the examples provided in the section 5.3 of the analysis, certain assumptions can be made. These words could indicate utterances such as: sources revealed, documents revealed, radaronline revealed exclusively etc.

In comparison, when examining the keyword list in the TMZ corpus, words such as trump, video, beckham, rocky, ring and died occur more frequently. This keyword list includes more names of celebrities than the RO keyword list. From the top keywords of the TMZ corpus, it is difficult to formulate assumptions of utterances or sentences without knowing more about the contexts, even though the names of the celebrities, such as Trump and Beckham provide some hints. The keyword lists are presented below.
Picture 2. Keyword list showing words that are significantly more frequent in the RO corpus than TMZ corpus.

Picture 3. Keyword list showing words that are significantly more frequent in the TMZ corpus than RO corpus.
2) Concordance, Collocates and Clusters

The Concordance -tool in AntConc shows how words and phrases are commonly used in the corpora and what kind of context they appear in. For investigating concordances any desired search terms have to be chosen by the researcher. The selected search terms are then inserted in the software, which executes the investigation (Anthony, 2014).

The selected search terms in this study are some of the keywords and most frequently occurring words that were investigated and discussed above: star/stars, trump and source/sources. What is more, additional search terms are included and chosen according to certain re-occurring themes of the celebrity news articles, which deal with the following topics: law suits, relationships and break-ups.

The Concordance – tool shows the context where and how the search term is used in the corpus, as well as the amount of concordance hits, that is, how many times the selected search terms appears in the corpus (frequency). The concordance hit search results are presented in the table underneath.

Table 6. The concordance hits of the selected search terms.

<table>
<thead>
<tr>
<th>Search Term</th>
<th>RO Concordance Hits</th>
<th>TMZ Concordance Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>star/stars</td>
<td>90/5</td>
<td>18/3</td>
</tr>
<tr>
<td>source/sources</td>
<td>69/12</td>
<td>3/7</td>
</tr>
<tr>
<td>reality</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>celebrity/celeb</td>
<td>0/0</td>
<td>0/2</td>
</tr>
<tr>
<td>breakup/split</td>
<td>3/12</td>
<td>2/5</td>
</tr>
<tr>
<td>boyfriend/girlfriend</td>
<td>8/14</td>
<td>7/1</td>
</tr>
<tr>
<td>documents/docs</td>
<td>31/3</td>
<td>0/14</td>
</tr>
<tr>
<td>court</td>
<td>54</td>
<td>13</td>
</tr>
<tr>
<td>show</td>
<td>60</td>
<td>11</td>
</tr>
<tr>
<td>divorce/wedding</td>
<td>25/28</td>
<td>10/1</td>
</tr>
</tbody>
</table>

When examining the search term star, the RO corpus has 90 hits. Even though there are 90 articles in the corpus, it is not one concordance hit per article though; some articles have the search term more than once, some do not mention it all. When inserting the term in plural
stars, there are only 5 hits. The usage of the same search word (both singular and plural) have remarkably less hits in the TMZ corpus.

The investigation shows, however, that the word celebrity and its abbreviation celeb do not occur at all in the RO corpus, and the TMZ corpus has the word celeb twice. Instead, the word star was relatively frequently used when talking about celebrities. It can also be seen from the table that the word trump is used more frequently in the TMZ corpus.

Furthermore, some of the most frequently occurring words from the corpora (star, trump) and the words celebrity/celeb are checked against COCA to see if there are similar patterns in their frequencies. The examination is conducted by calculating percentage frequencies. These results are presented in the table below.

**Table 7.** Percentage frequencies of some of the most frequently occurring words (star, trump) and the words celebrity/celeb that are extremely rare in the specialized corpora, and their equivalencies in the reference corpus.

<table>
<thead>
<tr>
<th>Word</th>
<th>% frequency RO</th>
<th>% frequency TMZ</th>
<th>% frequency COCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star</td>
<td>0,29</td>
<td>0,12</td>
<td>0,01</td>
</tr>
<tr>
<td>Stars</td>
<td>0,02</td>
<td>0,02</td>
<td>0,01</td>
</tr>
<tr>
<td>Celebrity</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00*</td>
</tr>
<tr>
<td>Celeb</td>
<td>0,00</td>
<td>0,01</td>
<td>0,00*</td>
</tr>
<tr>
<td>Trump</td>
<td>0,05</td>
<td>0,23</td>
<td>0,01</td>
</tr>
</tbody>
</table>

*Although the words celebrity and celeb are marked as 0,00 % in the table, they do occur in the COCA corpus (their concordance hits are 11102 & 461), however, since the dimension of the corpus is very large (around 570 million words), the percentages are rounded up to the nearest second decimal, the final percentages become 0,00.

As the table shows, the usage of star is higher in the RO and TMZ than in the COCA corpus. However, the percentage frequencies of the selected search terms follow similar patterns among all the corpora, as the differences in the percentage frequencies are not significant.

Out of the chosen search terms above, the word star is selected for further analysis for its collocates and clusters, and how it is used in the texts. The word is chosen as it has one of the highest concordance hits, and seems to be a remarkable word for the investigation, as it has replaced the word celebrity almost entirely in the corpora. The search term star is only investigated in its singular form.

**Step by Step: Further Analysis of the Selected Search Term Star**

Firstly, as mentioned previously, the Concordance–tool in AntConc shows how the word is used in context, therefore, it can be used to uncover how the word is used in real texts in the corpora. The pictures below demonstrate the usage of star, what kinds of associations it has
and what kinds of contexts it appears in. The most important words in the context are highlighted with different colors.

Picture 4. The search term star used in context in the RO corpus by using the Concordance-tool.

As picture 4 shows, in the RO corpus the word star is often related to types of celebrities, even though the word celebrity itself is not used. The types of stars (or celebrities) that are most commonly mentioned are stars from different reality TV shows and other television programs.
As it can be seen from picture 5, in the TMZ corpus the search term *star* is most frequently linked to porn stars, musicians, and reality TV show personalities, and other TV program stars.

Secondly, by using the Collocate–tool in AntConc, non-sequential patterns of *star* are examined. These are the top lexical collocates and the words that regularly appear near each other, and their relationship is statistically significant (Anthony, 2014). The table below summarizes these findings in the corpora.

**Table 8.** Summary of the top lexical collocates and frequencies of *star* in the corpora.

<table>
<thead>
<tr>
<th>Specialized Corpus</th>
<th>Top Lexical Collocates of the search term <em>star</em></th>
<th>Frequency of <em>star</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>RO</td>
<td>reality, former, porn, teen, mom, co, rhoc, real, rhony</td>
<td>90</td>
</tr>
<tr>
<td>TMZ</td>
<td>teen, mom, porn, spotted, sharp, ryan, rap, people nba</td>
<td>18</td>
</tr>
</tbody>
</table>
The results in table 8 confirm the previous findings of the concordance analysis. In the RO corpus, star co-occurs most frequently with reality, former, porn, teen, mom, co, rhoc, real, and rhony. Thus it can be stated that star is often used with types of reality TV shows (rhoc and rhony are initialisms that mean the reality TV shows Real Housewives of Orange County and Real Housewives of New York, and teen and mom refer to a reality TV show called Teen Mom), or when referring to an “ex” celebrity: former star, or starring something together with someone (co-star). Even porn industry is represented on the list.

The top lexical collocates of star in the TMZ corpus confirm the findings of the Concordance analysis as well. Star frequently co-occurs in context where it refers to sports (nba), porn industry, reality TV (Teen Mom), and music (rapper). From picture 5 it can be noted that ryan refers to Ryan Edwards, who is a reality TV personality from Teen Mom (Internet Movie Database), and sharp refers to Keesha Sharp, who is an actress, director and writer (Internet Movie Database).

Thirdly, Clusters –tool is applied next in AntConc in order to search for words (or patterns and groups) that result together immediately to the left or right of the search term star. This tool summarizes the results that were generated in the previous parts of the analysis (Anthony, 2014). In AntConc, it can be selected how many words the Clusters- tool shows appearing with the search term. In this study, the setting is set to show five words from left and right of the search term in order to get a full understanding of the word groups and patterns. The pictures below provide the word groups and patterns from left and right of the word star in both corpora.
It can be seen from the clusters analysis that in the RO corpus when examining the word patterns from the left of the search term star, the most results show a case where a name appears together with the search term (for example, star Alyson Stoner, star Aubrey O’day), or star is linked together with someone or something by using the word and (for example, star and her estranged husband, star and backing dancer revealed). When examining the
cluster from right of the search term, the results are more linked to a profession, or a type of star (for example, co-star, porn star, pop star, nba star).

![Picture 8](image1.png) **Picture 8.** The clusters from left of *star* in the *TMZ* corpus.

![Picture 9](image2.png) **Picture 9.** The clusters from right of *star* in the *TMZ* corpus.

It can be noted from pictures 8 and 9 that the clusters analyses follow the same patter in the *TMZ* corpus as it did in the *RO* corpus. When examining the clusters from left of the search term, it often appears together with names (for example *star Althea Eaton, star Keesha*
Sharply), while on the right of the search term different types of stars are shown, for example *people v. o.j star, teen mom star* and *rap star.*

The reference corpus is also investigated in order to compare its results with the RO and TMZ corpora. It has to be taken into consideration though, that in COCA not all *star* words relate necessarily to celebrities. The usage of *star* is presented in its contexts in the pictures below.

![Picture 10](image-url). The search term *star* used in context in the COCA corpus (concordance hits ranked from 1 to 16).

The picture shows that the first concordance hits of the usage of *star* refer often to newspapers or magazines that have the word *star* in their name. However, after examining the next concordance hits in COCA for *star,* the results show celebrity related contexts as well. These are presented in the picture underneath.
Picture 11. The search term *star* used in context in the COCA corpus (concordance hits ranked from 23 to 41).

### 6 Discussion

#### 6.1 Results of this Study

This study has examined culturally U.S. based celebrity news websites with the aim of identifying how journalists use language when reporting on celebrity. The study has investigated linguistic phenomena, patterns, structures and stylistic features that are typical for celebrity news language usage.

Identifying the usage of these linguistic phenomena and their subcategories in the celebrity news articles indicated that the websites have different writing styles, even though the topics of the websites and articles are similar (occasionally even the same news report on the same celebrity or incident). The results presented above indicated that the linguistic phenomena that are the most typical for the celebrity news coverage are found in the area of lexicon and vocabulary. This finding applied to both *Radar Online* and *TMZ*. Both websites contain unconventional spelling forms, colloquial expressions and slang, as well as taboo language. Within these linguistic subcategories, the analysis showed that there is more taboo language (which is also more vulgar and profane) in *Radar Online* than *TMZ*. When it comes to colloquial expressions and slang, the websites mainly used different kinds of words and expressions, but *TMZ* employed slang and colloquialisms more often than *Radar Online*. 
Word blending, g-dropping and abbreviations were also used in both Radar Online and TMZ, although these aspects were more typical for TMZ as well.

Nevertheless, the analysis showed that even though the celebrity news articles share similar characteristics, some phenomena are more frequently present in one celebrity news website than in the other. For example, as the results designated, TMZ contains more non-standard syntax and grammar structures, while Radar Online follows more general language guidelines. However, as it can be seen from the examples of the rhetorical devices and critical linguistics aspects, both websites use fragmented and simple sentences, for example rhetorical questions such as What? Dated? and The only problem?

Moreover, the subcategories of phonology were more frequently used in TMZ than Radar Online, as well as unconventional spelling forms. Thus it could be argued, that despite the fact that taboo language is more frequent and vulgar in Radar Online, overall the language usage of Radar Online is more conventional and closer to Standard English norms than the language usage of TMZ.

In addition, the results of the analyses showed that both celebrity news websites contain a lot of face-to-face spoken conversation features. For example, the websites are mixing colloquialisms, slang and formal words, which functions as a way to make the language more interesting for the target readers, bringing face-to-face conversational elements to the writing. The unconventional spelling forms that the authors are using function the same way: they are trying to imitate speech in the writing (Herring, Stein & Virtanen, 2013).

In addition, Conboy (2014) argues that this special kind of vocabulary of celebrity news indicates that the news outlet is talking to its readers in its own informal manner. This implication of this language use and special lexical choices are on the side of the people as readers, as opposed to the interests of the power-bloc. This can be seen as a deliberate strategy to establish the ideological bond with the readers (Conboy, 2014). What is more, this can also be seen as a rhetorical device, as by reaching out to the readers and having “a dialogue” with them, the writers are persuading the readers to agree with their message.

As the results of the analyses also demonstrated, both websites address to the reader and themselves (the website) repeatedly in their writing, which could be seen as an additional way to make the writing resemble speech (Herring, Stein & Virtanen, 2013). Also, the usage of pronouns in the articles is another rhetorical device, and it functions the same way as having a dialogue with the readers: the writers are trying to pursue them to agree with what they are saying.
Another aspect that brings conversational elements to the articles is the way the authors use irony and add their own comments of the events they are reporting on. These features not only bring spoken language elements to the articles, but the writers are also being subjective about their writing, employing playful tone and expressing their attitudes towards the subject (for example writing comments such as too bad, interestingly, oddly specific). The tone and irony in the celebrity news articles are presented in a humoristic way.

Also, the language usage is generally very simplistic and vernacular in the articles as well, for example consisting of sentence fragments (as stated before), simple grammar and simpleminded and unsophisticated vocabulary. The simplistic language in the articles is trying to resemble the language use of “everyday life”. These lexical choices can be seen as another feature of spoken communication.

The results of the corpus linguistics analysis revealed that celebrities are mainly constructed by referring to them as stars. This rule applied to all cases except two instances of TMZ where they had used the abbreviation celeb instead. This is also one of the reasons why star was selected for further analysis (as well as the fact that it had one of the highest concordance hits in the corpora). In addition, star was analyzed more thoroughly in order to discover, how celebrity is constructed linguistically in celebrity journalism. The word celebrity was completely absent in the corpora. The fact that the word celeb was used (although only twice in the corpora) instead of celebrity, refers to the slang and abbreviation usage, that seems to be typical for celebrity news coverage.

According to Baker (2006), there are usually exceptions to any patterns or rule, and in some cases what is not said is more important than what is actually said. The absence of celebrity and replacing it with the word star can be seen as a significant finding, as modern celebrity has previously been divided into two main categories: the film star and the television personality (Conboy, 2014). Based on the corpora analyses of this study, star applies to all kinds of celebrities in the contemporary celebrity culture and celebrity journalism. This finding is further confirmed by analyzing the top lexical collocates that frequently co-occur with star. Based on the findings of the analyses, the most common associations with star in the corpora are: reality TV show star, former star, porn star, co-star, rap star and NBA star. Thus, it can be noted that all types of celebrities are called stars, whether they are achieved celebrities (for example rap star and NBA star), attributed celebrities (for example porn star) or celletoids (for example reality TV stars). The remark that reality TV show stars had one of the highest occurrences in the corpora can be seen as an indicator of the strong influence of
celebrification, the making of a celebrity, which also represents the shift of “ordinary” people becoming famous without any particular talent.

The associations mentioned above are validated by the results of the examination of the word groups and word patterns that star most frequently co-occurs with. The analyses of the word clusters confirmed the previous findings that star co-occurs either with the type of star, (how or why that person is famous), together with a name of the person, or together with another person (for example star nikki benz, star and close pal).

When comparing the results to the reference corpus COCA, similar patterns were found in the frequency percentages of the search terms, as the differences between the frequency percentages were not significant. When comparing the usage of star in real life contexts in COCA, the top results showed cases where star referred to other concepts than celebrity. However, despite the fact that star as a celebrity were not the first rankings in COCA, celebrity related contexts with star were still high in the ranking, demonstrating that star is a remarkable association of celebrity.

When trying out different search terms in AntConc and examining their concordance hits, the search terms were chosen according to the most popular topics in the celebrity news articles. These were, for example, legal problems and relationship issues, such as divorce and break-up. Radar Online had more concordance hits in all these areas. Trump, however, was a search term that had more concordance hits in TMZ than Radar Online. Trump was also a word that had one of the highest concordance hits in the TMZ corpus. This could indicate that Donald Trump is more frequently discussed in TMZ. On the other hand, it has to be taken into account that not all Trump related words necessarily refer to Donald Trump himself, but to other members of his family as well.

All in all, it could briefly be summarized that when reporting on celebrity, journalists use very simplistic and vernacular language, having their own special kind of dialogue with their readers. The reporting consists of conversational elements that resemble spoken features, such as slang, colloquialisms, taboo language, simple grammar, onomatopoeia, irony and humor. By this special kind of language use, the writers are trying to control the way the readers comprehend and receive their message. The linguistic construction of celebrity follows these phenomena, and it is conducted by the word star.

6.2 Celebrity Journalism – Hybrid Journalism?
In popular and academic debate, celebrity journalism and celebrity news have been accused of dumbing down the news content and quality, with clichéd topics, inaccurate journalism
standards, colorful language and poor source verification. What is more, celebrity journalism has claimed to distract audience’s attention away from “more important issues” (Van den Bulck, Paulussen & Bels, 2017). However, Dubied & Hanitzsch (2014) point out that all the criticism and issues that are generally linked to celebrity journalism may as well apply to the more traditional forms of journalism as well.

Connell (as cited in Conboy, 2014, p. 173) has stated that “celebrity coverage acts as a commentary on the disparity between social classes and the resultant gulf in wealth in western societies”, adding that “such observation does not find an expression elsewhere in our mediascape”. Connell concludes that the popular tabloids have generated a necessary, if limited awareness of these discrepancies (as cited in Conboy, 2014, p. 173). Adding to Connell’s insight, Conboy (2014) continues that celebrity has diversified to act as a prismatic panorama on most aspects of contemporary discourses, such as gender, politics, mental health, sexual morality and national identity.

According to Conboy (2014), much has been written about the trend towards including more celebrity-based news in all media formats in general, and tabloids are leading this trend towards a more generalized tabloid culture, compounding it with a language style that matches that tabloid culture. Tabloids not only report on celebrities, but they also use that information to develop as many news stories as possible with a plethora of intertextual references to them. In a way, this could be seen as “democratizing” the news, as it moves away from its traditional insistence on the elites of society and a preponderance of financial and political reports. On the other hand, it might also limit the reach of news agenda and restrict the people who are considered as the “elite” (Conboy, 2014).

Turner (2014) argues that celebrity news really is news, “but not the news as we used to know it” (p.151). The lower status of celebrity news has to do with its compromised relation to more traditional version of news, as well as its implication in some of the changes that occur in how the news is understood as an informational and cultural form. Celebrity news is a product and appreciation of news journalism serving the interest of those with an industrial stake in publicizing the celebrity (Turner, 2014).

Furthermore, celebrity news has adopted several practices and discourses of “traditional” news, while successfully colonizing parts (or all) of many of the platforms that were once occupied only by other versions of news. It has to be noted that celebrity news satisfies numerous of the same cultural demands as other news (with gossip and fact overlapping), serving very different functions in terms of informing the audience to those traditionally attributed to journalism. Celebrity news has its own special way of industrial
production that is tightly manufactured for the needs of entertainment industries. In effect, celebrity journalism might be the first genre of news, which sole objective is to entertain the audience (Turner, 2014).

Van den Bulck, Paulussen & Bels (2017) suggest that celebrity journalism should be seen as a hybrid journalistic genre, rather than merely classifying it the “dark side” of a dichotomy between good and bad journalism. This would help to create public and academic recognition for a branch of journalism that constitutes a remarkable section of all news media, and allow understanding of the rich complexity and variability of journalism genres that goes beyond the boundaries of celebrity journalism as well (Van den Bulck, Paulussen & Bels, 2017).

7 Conclusions
Two U.S. based online celebrity news websites, Radar Online and TMZ, have been examined in this study with the aim of investigating how journalists use language when reporting on celebrity. The study has been concerned with how celebrity is constructed linguistically, therefore the study has attempted to highlight celebrity journalism as a specific genre, and how a specific type of language is required for mediated representations of celebrity.

This study has been conducted by using qualitative and quantitative research methods and approaches. Identification of linguistic phenomena, critical linguistics, rhetorical analysis and corpus linguistics have been applied in the study. The study has focused only on very recent celebrity coverage, as the timeframe that was followed covered only one month, March-April 2018.

After the analyses, the study was able to find answers to the research questions posed in the introduction:

RQ1: The study identified that the typical linguistic phenomena, features and structures used in the Radar Online and TMZ celebrity news articles are: 1) special kind of vocabulary and lexicon (consisting of abbreviations, slang, non-standard spelling, taboo language and colloquialisms), 2) non-standard grammar and syntax structures, 3) conversational spoken language elements that contain humoristic and ironic components, as well as 4) simplistic sentence structures. Another typical feature in the celebrity news articles is 5) subjective writing, that is, the writers of the articles often have a playful tone in their writing, indicating
that they do not take the subject-matter too seriously, whilst expressing their own attitudes towards the subjects.

**RQ2:** The analyses showed that the most frequently used linguistic phenomenon in both the celebrity news websites is in the area of vocabulary and lexicon. This category was further divided into subcategories of unconventional spelling, colloquialisms and slang, and taboo language. The particular kind of vernacular vocabulary that is typical for celebrity journalism is linked to everyday life, and it is applied in the celebrity news coverage in order to have a special kind of dialogue with the readers.

**RQ3:** By examining the characteristics of the celebrity news articles, it can be noticed that while the two celebrity news outlets share the same linguistic phenomena and features, some of the phenomena are more frequently used in one celebrity news outlet than in the other. For example, *TMZ* uses more non-standard grammar structure and unconventional spelling forms than *Radar Online*, as well as slang, colloquialisms and onomatopoeia. On the other hand, *Radar Online* includes more taboo language, and its usage is more offensive and vulgar compared to the taboo language of *TMZ*. However, it could be argued that overall the language usage of *Radar Online* is more restrained, moderate and closer to Standard English norms, as its style is not as free-spoken and crude as the style of *TMZ*. In addition, the celebrity news outlets differ from each other in terms of length of the articles and stylistic features, as the *TMZ* articles include more humoristic and ironic comments than the *Radar Online* articles.

The linguistic construction of celebrity consists of a vernacular vocabulary, simplified grammar and sentence structures, as well as conversational features in the text, such as irony, humor and addressing directly to the reader. In addition, celebrity is mainly constructed by the usage of the word *star*, and this applies to all kinds of celebrities: reality TV personas, musicians, actors, athletes, and even to the industry of pornography. The word *celebrity* itself, or its abbreviation *celeb* are almost non-existent when reporting on celebrity.

The high frequency of reality TV related stars demonstrates the popularity of reality TV formats, and it shows the strong influence of celebification in the contemporary celebrity culture and entertainment industries.

Celebrity journalism and celebrity news have been accused of dumbing down the news content and quality, even though all the criticism and problems that are usually
linked to celebrity journalism may as well apply to the more traditional forms of journalism as well. Celebrity journalism could be seen as a hybrid journalistic genre, rather than just classifying it the “dark side” of a dichotomy between good and bad journalism. Celebrity news can be classified as news, even though not necessarily in the same way as more traditional news, however, celebrity news has adopted several practices and discourses of traditional news, while serving very different functions in terms of informing the audience to those traditionally attributed to journalism. Much has been written about the trend towards including more celebrity-based news in all media formats in general, and tabloids are leading this trend towards a more generalized tabloid culture, compounding it with a language style that matches that tabloid culture.

Finally, it can be concluded that when reporting on celebrity, journalists use very simplistic and informal language, having their own special kind of dialogue with their readers. It consists of conversational elements that resemble spoken features such as slang, colloquialisms, taboo language, simple grammar, onomatopoeia, irony and humor. By this special kind of language use, the writers are trying to control the way the readers comprehend and receive their message. The linguistic construction of celebrity follows these same phenomena and patterns, and it is conducted by the word *star*.

The two celebrity websites that were investigated in this study represent the biggest celebrity and popular culture in the Western countries. Their media production is very active, and as online publications they also represent new celebrity news formats, therefore, their media production is also easily distributed and accessed 24 hours a day. Thus it could be stated that *Radar Online* and *TMZ* could be viewed as typical examples of celebrity news journalism.

It could be assumed that the linguistic phenomena and language patterns that were found in this study appear in other types of articles and texts in the genre of celebrity journalism and celebrity news as well. It could also be assumed that similar results could be identified in other celebrity news media from other countries and in other celebrity news articles that are written in different languages. However, in order to state this, more research on other celebrity news outlets is needed. Possible future research projects would be analyzing larger corpora than the ones that were used in this study, that is, adding more celebrity news outlets and increasing the number of articles for the investigation. What is more, different kind of methodology could be incorporated as well, for example conducting interviews of editorial work.
Additionally, different kinds of genres could be analyzed as well, and compared to the genre of celebrity news. For example, an analysis and comparison of how celebrity is reported on and constructed in more traditional hard news versus typical celebrity journalism could be an interesting viewpoint for the next research project.

8 References


**Appendix A) Radar Online Articles**


https://radaronline.com/exclusives/2018/03/stassi-kristen-katie-call-arianas-brother-creepy-ahead-date-billie-lee/


Appendix B) TMZ Articles  


http://www.tmz.com/2018/03/12/hubert-givenchy-dead/  


62

http://www.tmz.com/2018/03/30/donald-trump-spring-interns-white-house-white-out/


