

# “Hope you’re in the mood for 🍪ookies”: An Exploratory Study of Individual Writing Styles Across Social Media Platforms

Karoline Marko<sup>1</sup>, Grace Sullivan Buker<sup>2</sup>

<sup>1</sup>University of Graz, Austria

<sup>2</sup>Northeastern University, USA

---

## Article Info

### Article history:

Received April 17, 2022

Revised June 15, 2022

Accepted June 30, 2022

---

### Keywords:

Social media

Individual style

Computer-mediated communication

Case study

Authorship analysis

---

## ABSTRACT

In the past decade, research on computer-mediated communication (CMC) has flourished. The present paper examines social media posts taken from four users of Twitter and Instagram and aims at shedding light onto linguistic similarities and differences on the styles platforms. The focus is on platform-dependent features as well as on platform-independent features typical for each user (i.e., idiolectal features to be utilized in forensic authorship analysis). The dataset comprises 1.200 posts and is analyzed in terms of structural and CMC features, including sentence lengths and complexity, punctuation marks, emoji, hashtags, @mentions, and capitalization. The analysis shows that while the language used on Twitter does indeed differ from the language used on Instagram, some features are also potentially stable across the platforms, suggesting the existence of different registers. Further analyses will have to investigate a larger corpus to test the reliability of the results.

*This is an open access article under the [CC BY-SA](#) license.*



---

## Corresponding Author:

Karoline Marko

University of Graz, Austria

Email: karoline.marko@uni-graz.at

---

## 1. INTRODUCTION & BACKGROUND

In the past decades, research in the field of computer-mediated communication (CMC) has flourished, incorporating a variety of methodologies (see, e.g. Biber and Egbert, 2018; Herring, 2004; Tagg, 2015; van Dijck and Poell, 2013). For the purposes of this paper, discourse analytical and stylistic methods are applied to the analysis of social media posts made to the platforms of Instagram and Twitter. The data and analysis are approached from two angles: firstly, a discourse analytical point of view is taken with the aim of investigating differences across the two platforms that potentially suggest the emergence (or existence) of two platform-specific registers. Secondly, a forensic linguistic perspective is used to approach similarities in language use across the two platforms that indicate more platform-independent and thus potentially idiolectal language use.

Forensic linguistics (FL) is a field within linguistic analysis that analyzes data from a variety of legal contexts: for example, one of the tasks of forensic linguists relates to the comparison of two or more texts with the aim of proving or disproving common authorship (e.g., Coulthard, Johnson and Wright, 2017). Already back in 2012, MacLeod and Grant have suggested that “traditional methods [of authorship analysis] do not easily translate into computer-mediated communication” (p. 210), and also observe an “increasing use of online communications for the organization and dissemination of a wide range of criminal activities and material” (p. 201). CMC is largely perceived to have the potential of being anonymous (Herring and Stoerger, 2013), which makes such data particularly important to FL enquiries. Throughout the past decade, extensive but mostly computer-linguistic research has suggested that social factors such as gender and age (see, e.g., Herring and Stoerger, 2013; Marwick, 2013; Nguyen, Gravel, Trieschnigg and Meder, 2013; Peersman, Daelemans and van Vaerenbergh, 2011) can be extracted from online text. Grant’s (2010) work on language style in SMS text messages and Coulthard et al.’s (2017) investigation of authorship in a kidnapping case exemplifies possible applications of findings of CMC research to forensic contexts.

## 2. LITERATURE REVIEW

Since this paper addresses the language and discourse on social media platforms from two perspectives, the literature review is separated into two parts. Firstly, CMC will be discussed in connection to forensic linguistics, and subsequently findings from research on CMC features on social media platforms are presented.

### 2.1 CMC and Forensic Linguistics

Authorship analyses, particularly in CMC context, are not straight forward (e.g., MacLeod and Grant, 2012; Sousa Silva et al., 2011). The quality of an analysis depends, for instance, on the quality of the available material and texts lengths (Ehrhardt, 2018) – the latter being a particular issue with CMC texts. That is, texts produced digitally, such as text messages or social media posts, tend to be significantly shorter than other texts, such as letters, which are commonly encountered in forensic contexts, and thus prove more difficult to use for most forms of authorship analysis, particularly automated and statistical approaches (e.g., Coulthard et al., 2017; Grant, 2013). Even though several researchers have begun to address authorship problems in relation to digital technologies (e.g., Ishihara, 2017; Layton, Golder and Lotan, 2010; MacLeod and Grant, 2012; Orebaugh and Allnutt, 2009), to our knowledge, it has not yet been attempted to track an individual's style<sup>1</sup> across two social media platforms to investigate idiosyncratic linguistic features from a qualitative perspective. Thus, this study is the first of its kind to aim at detecting consistent (potentially platform-independent) and non-consistent (potentially platform-dependent) linguistic features of an individual across Twitter and Instagram. As outlined above, the investigation of platform-independent features can be applied in forensic contexts, while the analysis of platform-dependent features indicates the emergence of different writing styles on the investigated platforms. Thus, this research contributes equally to forensic linguistic and CMC research.

### 2.2 CMC and Social Media Platforms

Social media platforms have become extremely popular: in 2019, it was estimated that, world-wide, 3.48 billion people used such platforms regularly (Kemp, 2019). To date, much linguistic research has focused on Twitter (see, e.g., boyd, Golder and Lotan, 2010; Bruns and Moe, 2014; Dayter, 2015; Hardaker & McGlashan, 2016; Zappavigna, 2013), but little focus has been put on Instagram. While Instagram is primarily a picture-sharing platform with a focus on the visual (Leaver, Highfield and Abidin, 2020), Twitter offers a form of microblogging (i.e., blogs limited by length) and even though the sharing of pictures is possible, it is optional (Zappavigna, 2013). In contrast, Instagram allows the sharing of pictures without captions, while the sharing of captions alone is not possible (Leaver et al., 2020). Zappavigna (2013) identifies the following features as typical of microblogs: non-standard orthography, emotions (and by implication also emoji), hashtags<sup>2</sup>, abridged posts, address and @mentions<sup>3</sup>, and retweets<sup>4</sup>. Other features of digital language more generally are, for instance, the use of capital letters to indicate intense feelings, the repetition of letters for emphasis, and hyperbolic punctuation (McCulloch, 2019). These features are said to convey an additional layer of meaning which add a personal note to the text and facilitate the transfer of non-linguistic and paralinguistic signals, such as intonation. McCulloch (2019:153) even suggests that differences in the use of interpersonal features such as expressive typography “can tell the difference between writers” – a notion that is at the heart of forensic authorship analysis.

Being a picture-sharing platform, Instagram prioritizes visual imagery, which is a possible explanation for why so little linguistic research has addressed the language use on Instagram. The audience's participation in the shape of liking, commenting, or sharing, among others, is optional, asymmetrical to the user, and is an example of Jenkins et al.'s (2013) notion of “spreadable media” in which media is shared by a user for their own purposes. Given the plentitude of multi-modal material on Instagram, Brownlow, Pate, Alger and Naturile (2019) have analyzed the language of captions in posts by celebrities, and Veum and Undrum (2018) have focused on the language used in selfies posted to Instagram. They have found that the captions accompanying selfies can be compared to what Fairclough (1992, 1995) has termed “advertising style”, since it includes the use of personal language, lyrical elements (such as quotes or parts of song lyrics), and elements that address the reader directly.

Herring (2004:355) provides a useful framework for the analysis of CMC. Her framework includes the analysis of language on four levels, namely the structural level, the level of meaning, the interactional level, and the level of social behavior. Being the first analysis of its kind, the focus of this paper will be restricted to the structural features of language, leaving an investigation of higher levels of language (as suggested by Grant and MacLeod, 2020) to future research.

As boyd (2010) suggests, social networking services have similar attributes: persistence (capturing and archiving of content), replicability (duplication of content), scalability (broad visibility of content) and searchability (access to content via search). Paige, Barton, Unger and Zappavigna (2014) describe social media

---

<sup>1</sup> As “writing style” is an inherently difficult concept to analyze, we focus here on relatively quantifiable stylistic elements, such as the use of punctuation, and sentence lengths, as well as CMC-specific features such as the use of capitalization for emphasis, and the use of emoji, hashtags, and @mentions. For discussions of writing style, see, e.g., Krieg-Holz and Bülow (2016)

<sup>2</sup> A “form of metadata labelling the topic of a tweet” (Zappavigna, 2013, p. 83), marked with #

<sup>3</sup> The @ sign is used to address another user directly (Tagg, 2015)

<sup>4</sup> Indicates that “a message has been forwarded” (Paige, 2012, p.183)

platforms in terms of participant structure, audience engagement, purpose and activity, participant characteristics and self-disclosure, and media characteristics. At the beginning of our analysis (see Table 1 below), we compare Twitter and Instagram according to these features in order to point out platform-specific differences and similarities that can later help explain some of the findings presented in this paper. Motivated by the CMC and FL literature, the following research questions have been developed:

- For the selected population, are any of the investigated structural and/or CMC features stable (i.e., not influenced by the platform) across the social media platforms Twitter and Instagram? How can these findings be helpful to FL researchers and practitioners?
- For the selected population, are any of the investigated structural and/or CMC features influenced by the platforms (i.e. platform-specific)? How do these findings further our understanding of CMC on the investigated platforms?

### 3. DATA & METHODOLOGY

The data for this paper was collected manually from the respective social media platforms and based on a previous study (Marko, 2020), four individuals from a pre-existing corpus were randomly chosen to be included in the present study. Two of the individuals are male, two are female, and their ages are 32 (Person A), 34 (Person B), 35 (Person C), and 46 (Person D). Their most recent 150 posts made to Instagram and Twitter, i.e. 300 posts in total per individual, are included in the corpus. The data was collected during a period of several weeks in late summer of 2019. Depending on the frequency with which the individuals post, the time span from which their posts are taken ranges from several weeks to several months.

Once the corpus was collected, the involved researchers independently annotated the texts for structural features of writing style (see below) and later compared these features for each person. A top-down approach was used based on the features identified in the literature as typical of CMC and relevant to forensic linguistic inquiries. For each individual, a comparison was made between their posts on Instagram and their posts on Twitter to establish which features are consistent across both platforms, if any. The following features were selected for analysis:

- Punctuation (e.g., Chaski, 2001; McMenamin, 2002, 2010)
- Number representation (e.g., McMenamin, 2002; Sousa Silva et al., 2011)
- Sentence complexity, as measured with Readability Ease Scores<sup>5</sup> and sentence length (e.g., McMenamin, 2002; Sousa Silva et al., 2011)
- Capitalizations for emphasis (e.g., McCulloch, 2019)
- Other CMC-specific features, such as @mentions, #hashtags, and emoji (e.g., Sousa Silva et al., 2011; Zappavigna, 2013)

Even though additional features such as the use of contractions, word choice, formality and phrase structures were examined as well, space restrictions prevent us from reporting on all features and thus it was decided to focus on the features outlined above.

#### 3.1 Ethical Considerations

Ethical considerations are particularly crucial in relation to social media data (Townsend & Wallace, n.d.). The considerations undertaken in the present study relate mostly to data protection and anonymity of the involved individuals. To protect the involved individuals, they are referred to via pseudonyms and, to prevent reverse identification via search engines, whenever examples are given in this paper, the language of the posts is modified in a way that still allows us to illustrate our arguments (see Ayers et al., 2018).

### 4. ANALYSIS & RESULTS

If it can be established that the investigated platforms elicit different writing styles, even if based on their affordances, it becomes possible to determine which linguistic features are rather stable across the two platforms and allows us to draw tentative inferences about platform-dependent features. As a starting point, a comparison is drawn between the affordances of Twitter and Instagram, as Table 1 shows. The aim for this paper is to provide a foundational methodology, which can later be applied to a larger corpus of social media data.

Table 1 highlights differences between the two investigated platforms which are likely accountable for language differences that are identified in this study. To understand each subject's writing style, the features identified for each person will be reported and discussed below.

---

<sup>5</sup> Readability is a number calculated based on the Flesch-Kincaid test of Readability which calculates total words over total sentences and was originally developed to test the ease of legibility and comprehension of United States military documents in the 1970s. The Flesch-Kincaid formula is now a functionality of Microsoft Word and determines reading ease, where higher scores are easier to read and lower scores are more difficult. *Time* magazine is rated to have a score of 50 (Flesch, 2016), whereas a sixth grade writing level would have a score of around 80-90.

Table 1. Features of Instagram and Twitter (see Paige et al., 2014).

Feature	Twitter	Instagram
<b>Participation Structure</b>	Asymmetrical: User posts texts, or image + text, audience not required to respond	Asymmetrical: User posts image + text, audience not required to respond
<b>Audience Engagement</b>	Audience (or followers) respond to user post with link to an image or text, or “retweets”/“favorites”	Audience (or followers) respond with comments only, or “likes”
<b>Main Purpose and Activity</b>	Stream of consciousness thoughts, Responses/Reactions to other users’ tweets	Documentation by uploading an image plus a caption
<b>Participant Characteristics &amp; Self-Disclosure</b>	Mainly text-based, identity through expression of thoughts and opinions  Public account, but possibility to use pseudonyms	Identity performance through documentation of self (selfies), friends (group selfies), or components of one’s life  Private or public account; use of pseudonyms
<b>Media Characteristics</b>	Text	Image + text

#### 4.1 Person A

To begin the analysis, consider an overview of the features found in Person A’s data (Table 2).

Table 2. Overview of Person A’s linguistic and stylistic features.

Features	Twitter	Instagram
<b>Punctuation</b>	End-of-post punctuation: 32% (63% periods; 20% exclamation marks; 12% question marks; 4% colons)	End-of-post punctuation: 40% (66% periods; 21% exclamation marks; 11% question marks; 2% ...)
<b>Number Presentation</b>	Numerals, e.g., “3 years”: 58% Number words, e.g., “first time”: 41%	Numerals, e.g., “3 years”: 76% Number words, e.g., “first time”: 24%
<b>Readability and Sentence Length</b>	Readability ease (RE): 67 Sentence length: 11.9 words	Readability ease (RE): 65 Sentence length: 9.8 words
<b>Capitalizations for Emphasis</b>	CAPS for emphasis: 14	CAPS for emphasis: 39
<b>Instances of Emoji usage</b>	Emoji: 12	Emoji: 63
<b>CMC-specific features</b>	@mentions: 3	@mentions: 44
<b>#hashtags</b>	Hashtags: 0	Hashtags: 10

As Table 2 illustrates, Person A’s writing style shows both similarities and differences across the two platforms. Firstly, although the number of posts containing final punctuation marks differs slightly, Person A uses similar amounts of periods, exclamation and questions marks. Interestingly, Person A regularly ends their posts without any punctuation marks at all. Secondly, a difference emerges with regard to the use of numerals and number words: on Twitter, the use of numerals is less common than on Instagram, while number words are more common on Twitter. Further, even though the lengths of sentences differ slightly, with Instagram posts containing shorter sentences than Twitter, the RE scores for posts across the platforms is similar (67 for Twitter; 65 for Instagram), indicating a similar complexity of sentences.

Regarding the use of capital letters for emphasis, the use of which is illustrated below in examples (1) and (2) from Instagram and (3) and (4) from Twitter, a large quantitative difference emerges: while Person A uses capitalization for emphasis only 14 times on Twitter, they do the same 39 times on Instagram:

- (1) “...something that works on a sweater absolutely SLAYS here”
- (2) “You get an aide to do your hair AND free snacks.”
- (3) “I love the official ‘HE WAS HERE’ graffito.”
- (4) “There’s a DEPARTMENT for us?”

Another platform-dependent difference is found in the use of emoji: Person A uses only 12 emoji on Twitter, while they use 63 emoji on Instagram. Examples (5) and (6) from Instagram and (7) and (8) from Twitter illustrate the use of emoji:

- (5) “Hope you’re in the mood for 🍪 cookies because I 🍪eated some.”
- (6) “Outtake 📷 by @name”
- (7) “Eavesdropping on the computer of an elderly man next to me -- he googled ‘fun things to do today’ 😊”
- (8) 🍪🍪

Emoji can serve a variety of functions<sup>6</sup> (see, e.g., Marko, 2020; Danesi, 2016; Evans, 2017; Gawne & McCulloch, 2019; Schneebeli, 2017). In Person A's dataset, a complementary function is the most prevalent. That is, on Twitter, 58% of Person A's emoji uses are complementary, meaning that they provide a meta-comment to the text. On Instagram, the functions served by emoji are more varied, yet the most common function remains complementation (41%), followed by reinforcement (36.5%) and substitution (14%).

Further CMC-specific features used by Person A are @mentions and hashtags. Interestingly, both features are much more common in their Instagram than Twitter posts. In fact, Person A makes no use of hashtags on Twitter at all, and only three @mentions are to be found. Examples (9) - (10) show the use of these features on Instagram and examples (11) - (12) show their use on Twitter:

- (9) "Mary: 'Turn me into a shining sultry pomegranate seed of a woman.'... @name: 'We can do that.'"  
 (10) "#girlboss"  
 (11) "I contributed in a tiny way to that project by @name and the idea (which they came up with) is fucking genius."  
 (12) "hats off to @name"

It is an important feature to note that all @mentions on Twitter and 97% of them on Instagram are integrated into the running texts, thereby directly replacing the names of people, companies, or other institutions. This same usage is also found in Persons B, C, and D, as discussed later. In contrast to the use of @mentions, Person A's hashtags on Instagram are placed after their posts rather than within. Thus, they constitute elements outside the actual post and serve as a meta-classification of the respective posts. There is only one stand-alone hashtag (example 9) in the whole of Person A's dataset.

#### 4.2 Person B

As Table 3 shows, Person B uses end-of-post punctuation more frequently on Twitter than on Instagram, and in contrast to Person A employs end-of-post punctuation at least in half of their posts on Instagram and in even 85% of their posts on Twitter.

Table 3. Overview of Person B's linguistic and stylistic features.

Features	Twitter	Instagram
<b>Punctuation</b>	End-of-post punctuation: 85%	End-of-post punctuation: 51%
	"...": 31 instances	"..." 15 instances
<b>Number Presentation</b>	Numerals, e.g., "3 years": 53%	Numerals, e.g., "3 years": 38%
	Number words, e.g., "first time": 48%	Number words, e.g., "first time": 62%
<b>Readability and Sentence Length</b>	Flesch Reading Ease: 69	Flesch Reading Ease: 65
	Sentence length: 11.8 words	Sentence length: 12.1 words
<b>Capitalizations for Emphasis</b>	CAPS for emphasis: 4	CAPS for emphasis: 12
<b>Instances of Emoji usage</b>	Emoji: 90	Emoji: 183
<b>CMC-specific features</b>	@mentions: 132	@mentions: 107
<b>@mentions</b>		
<b>#hashtags</b>	Hashtags: 166	Hashtags: 578

An interesting feature regarding Person B's writing style relates to their use of three dots as post-final punctuation (but also within posts), which can be observed on both platforms (see examples (13) - (16) below):

- (13) "Ireland... you're amazing. IE 🇮🇪"  
 (14) "I left my ❤️ in San Francisco..."  
 (15) "Food for thought... @name @name"  
 (16) "On another planet...#safari #lions #queenelizabethpark #uganda #africa #adventure #travel #nationalgeographic"

A highly interesting feature of Person B's writing style is their use (or lack) of hyphens. Examples (17) - (20) from Instagram and (21) - (23) from Twitter serve as illustrations.

- (17) "I'm thinking of starting a coed bobsled team"  
 (18) "coparenting day with @name..."

<sup>6</sup> Substitution: using an emoji instead of a word within a sentence; Reinforcement/emphasis: usually illustrative - emoji conveys that same idea as the accompanying words do; Contradiction: emoji use contradicts the meaning of the text, usually for ironic purposes; Complementation: meta-comment that adds an additional thought or idea that is not conveyed by the accompanying words

- (19) “please donate to make this ambitious, game-changing projects happen (and of course, is tax-deductible).”  
(20) “Always a pleasure to share post Olympic life advice”  
(21) “Pre interview pitstop for [name]’s favorite cookies. @name @name @name”  
(22) “Eye opening conversation with [name] tonight about their experiences.”  
(23) “What an interesting, in depth conversation”

These examples illustrate that there seems to be a particular pattern for Person B’s use of hyphens: a tendency appears to not use hyphens in words with the prefix “co” (spelled as one word), while words containing the prefixes “pre” and “post” are spelled as two words. Adjectival compounds do not seem to follow such a neat pattern, as they are found in both hyphenated and non-hyphenated forms.

In contrast to Person A, Person B uses more numerals on Twitter and more number words on Instagram. However, their RE scores are also rather stable on both platforms, which suggests a similar complexity of sentences. For Person B, sentence lengths differ only slightly across the two platforms. The use of capitalization for emphasis is lower on Twitter than on Instagram, in line with the findings for Person A.

Examples (24) - (27) illustrate Person B’s use of emoji. They appear to use emoji differently on Twitter than on Instagram, and, particularly, to a vastly different amount, as indicated in Table 3.

- (24) “Wedding halftime. 🍷 😊 📷: @name”  
(25) “Well said 🤝”  
(26) “👊”  
(27) “Happy belated #Earthday 🌍”

Regarding emoji functions, Person B uses predominantly emoji with a complementary function (55%), followed by a reinforcing function (36%) on Instagram; similarly, 46% of emoji on Twitter serve a complementary function and 31% serve a reinforcing function. Interestingly, 18.7% of Person B’s emoji are stand-alone emoji on Twitter, while only 3% of their Instagram posts contain emoji used in this way.

A further striking feature is Person B’s use of hashtags and @mentions. Firstly, they use 132 @mentions and 166 hashtags on Twitter and 107 @mentions and a total of 578 hashtags on Instagram. As examples (28) - (31) show, @mentions are used within sentences to replace the actual names of people and places. Importantly though, hashtags, too, are frequently integrated into sentences, but are also found to be placed outside post boundaries.

- (28) “thanks to @name for the important work you continue to do and on honoring these incredible human beings and #Olympic athletes. #inpursuitofpeace#internationalcrisisgroup#olympicrefugeeteam #refugees”  
(29) “A mini Rome 🏰 #annecy #france #wanderlust”  
(30) “Very true @name. Great article.”  
(31) “I think I found my inner #countrygirl. 🏠 📖 🚗 #blackrockdesert #deepplaya #burningman @Burning Man”

#### 4.3 Person C

As can be seen in Table 4, Person C’s writing style illustrates a variety of similarities and differences across the two social media platforms, many of which continue to represent the patterns found in Persons A and B.

Table 4. Overview of Person C’s linguistic and stylistic features.

Features	Twitter	Instagram
<b>Punctuation</b>	End-of-post punctuation: 16% (42% period; 33% question marks; 25% exclamation marks)	End-of-post punctuation: 17% (48% period; 12% question marks; 28% exclamation marks ; 8% ellipsis)
<b>Number Presentation</b>	Numerals, e.g., “3 years”: 73% Number words, e.g., “first time”: 27%	Numerals, e.g., “3 years”: 85% Number words, e.g., “first time”: 15%
<b>Readability and Sentence Length</b>	Readability ease: 63.9 Sentence length: 21.8	Readability ease: 65.7 Sentence length: 13.5
<b>Capitalizations for Emphasis</b>	3	39
<b>Instances of Emoji usage</b>	5	39
<b>CMC-specific features</b>	@mentions: 17	@mentions: 32
<b>@mentions</b>		
<b>#hashtags</b>	Hashtags: 3	Hashtags: 21

Some similarities can be seen in the use of punctuation: almost the exact same amounts of full stops (periods) at the end of their posts can be found (16% on Twitter; 17% on Instagram). In contrast to Person B, they prefer to end posts with no punctuation marks (even less so than Person A), but when there is a punctuation mark, it is more likely to be a full stop than anything else.

Regarding the presentation of numbers, Person C, in contrast to Persons A and B, prefers the numerical presentation of numbers on both platforms, even though more of them are used on Instagram. RE scores are, in line with previous findings, similar on both platforms, suggesting a similar complexity of sentences. Sentence lengths, however, differ more for Person C than for the other three individuals: with an average of 21.8 words per sentence, Person C's sentences are longer on Twitter than on Instagram, where the average length is only 13.5 words. Example (32) shows the sophisticated and complex sentence structures Person C uses, with complex NPs, introductory clauses, and rhetorical strategies:

(32) "While we're on this interesting topic, I also imagined that Beebe's Bathysphere was a kind of analogue to Plato's Retreat, not, as it turned out, a tomato-shaped submarine"

A discursive feature that varies across the platforms for Person C is their use of capitalization for emphasis, as well as their use of emoji. Person C very rarely capitalizes for emphasis on Twitter (3 instances), but does so very frequently on Instagram (39 instances). Similarly, the use of emoji on Instagram (39 instances) is much more common than on Twitter (5 instances). Regarding emoji, the dominant function Person C uses emoji for is complementation (Instagram: 69%; Twitter: 80%), but they also use reinforcement (21%) on Instagram. This, however, is not the case on Twitter, where stand-alone emoji (20%) are the second type of emoji used. Importantly for authorship analysis, the variety of functions in which emoji are used is limited to two per platform. Examples (33) - (36) illustrate Person C's use of emoji (the first two are taken from Twitter, the latter from Instagram):

(33) "👤 👤 VERY IMPORTANT NEWS 👤 👤"

(34) "😬"

(35) "🌰 nuts"

(36) "I wrote a thing about slacker prince [name] 🌸"

Similar to Persons A and B above, the use of hashtags for Person C is much more common on Instagram than on Twitter, as is their use of @mentions. In line with previous findings, the integration of @mentions into the running text is apparent. There are only ten instances of @mentions that are placed outside post boundaries and are not integrated into the context of the post as referential to the user being discussed. Hashtags, however, are not integrated into the sentences but are placed outside them.

#### 4.4 Person D

Finally, as the patterns continue to emerge regarding style and platform-specific usage, consider the last user, Person D.

Table 5. Overview of Person D's linguistic and stylistic features.

Features	Twitter	Instagram
<b>Punctuation</b>	End-of-post punctuation: 35% (73% period; 13% question marks; 10% exclamation marks)	End-of-post punctuation: 54% (95% period; 1% question marks; 20% exclamation marks)
<b>Number Presentation</b>	Numerals, e.g., "3 years": 73% Number words, e.g., "first time": 24%	Numerals, e.g., "3 years": 60% Number words, e.g., "first time": 43%
<b>Readability and Sentence Length</b>	Readability ease: 74.8 Sentence length: average 14.44	Readability ease: 65.7 Sentence length: average 13.2
<b>Capitalizations for Emphasis</b>	10	1
<b>Instances of Emoji usage</b>	16	138
<b>Medium-specific features</b>		
@mentions	@mentions: 26	@mentions: 45
#hashtags	Hashtags: 2	Hashtags: 56

As Table 5 shows, Person D's writing style across the platforms illustrates, again, similarities and differences. An interesting difference is seen with punctuation: they utilize punctuation at the end of their posts 35% of the time on Twitter, but 54% of the time on Instagram. For both platforms, periods (full stops) are the favored end-of-post punctuation, occurring 73% of the time on Twitter and even 95% of the time on Instagram. Furthermore, the user deviates their usage of question and exclamation marks at the end of their posts on both platforms: on Twitter, the use of question marks (and thus the involvement of the audience) is higher than on Instagram, while the use of exclamation marks is higher on Instagram than on Twitter.

The use of numerals to represent numbers is very high on both platforms, but, in contrast to the other three subjects, is higher on Twitter than on Instagram. Also, for the first time, RE scores differ by more than four points, which implies a larger variety of sentence structures compared to Persons A-C and a more complex sentence structure on Instagram. Person D writes with an average of 14.4 words for their posts on Twitter and of 13.2 words on Instagram.

Capitalizations for emphasis are not a feature Person D commonly uses; however, they use emphatic capitalization more commonly on Twitter (10 instances) than on Instagram (1 instance). Example (37) from Twitter and (38) from Instagram show two instances in which emphatic capitalization is used and in which the entirety of the post is capitalized:

(37) “ALT HED: RICH STILL EXTREMELY RICH UNDER WARREN TAX PLAN”

(38) “SUMMER ISN’T OVER”

This type of capitalization occurs ten times more frequently on Twitter than on Instagram, which may lend support to the theory that certain colloquial or stylistic features found in CMC are more prevalent on Twitter than Instagram, as it may pattern more closely to instant messaging genres than the blogging style that Instagram patterns more closely to.

Regarding the use of emoji, the functions with which they are used patterns differently from the other three users. In fact, the most frequent function for Person D on Instagram is reinforcement (65%), followed by complementation (31%). On Twitter, however, 50% of emoji serve a complementary function and only 19% serve as reinforcement. Their overall use of emoji is, as for Persons A-C, higher on Instagram than on Twitter. Examples (39) - (41) from Twitter and (42) - (44) from Instagram depict Person D’s use of emoji:

(39) “Dismantling capitalism one piece at a time. 🤖”

(40) “👁👁👁 [direct quote]”

(41) “When your kid is telling you how tired she is 🤖”

(42) “Dinner with @[name] and a martini 🍸😎”

(43) “Date night ❤️”

(44) “Rainbows. 🌈”

In terms of CMC-specific features, Person D’s behavior shows some of the same patterns found in the other three subjects: they utilize @mentions and hashtags more frequently on Instagram than on Twitter. Further, Person D integrates @mentions within the text in a seamless manner, as seen with Persons A, B and C, as shown in example (45) from Instagram:

(45) “At @contrany x @pujolrestaurant: 1,834-day old mole madre [omitted]”

This post starts with “at” to suggest the implicit meaning “I am at...” followed by a @mention. In this respect it is worth noting that the @-sign of the @mention appears to be invisible, unremarkable, or seen as a required feature to tag someone. Otherwise, the text would be read as “At at contrany” – that is, the initial “at” is reduplicative if the @ sign was read at “at” as well. Such a use of @mentions is worth investigating more closely in the future.

In terms of hashtag use, Person D shows an asymmetry. On Twitter, they only use two hashtags, but on Instagram they utilize 56 of them. The hashtags on Instagram are more summative of the respective posts, ostensibly providing details about the content of the image that goes along with the posts, or are included to notify other users who follow certain hashtags of their posts. Examples (46) and (47) from Instagram illustrate Person D’s use of hashtags:

(46) “Hydroponic red-veined sorrel at [name] outside San Juan. [name] PR #puertorico”

(47) “Tonight’s #sunset 🌅”

## 5. DISCUSSION

Before moving into a discussion of the findings, the small size of the corpus needs to be acknowledged. Since the nature of this study is exploratory and involves a large amount of manual tagging, the endeavor to build a larger corpus is left to future research. Furthermore, future research will need to investigate individuals from different age groups and with different social backgrounds to allow for a more comprehensive and complete picture to be drawn. Another possible limitation is the effect of the expected audience: as described in Table 1, it is more common for Twitter than for Instagram posts to be direct to specific individuals, and it is therefore acknowledged that the audience might have affected the language used in the posts (Bell, 1984; Goffman, 1981). Additionally, the Twitter user interface prompts its users to answer the question “What’s happening?” using short texts (i.e. tweets) that reach a network of other Twitter users (Marwick and boyd, 2010). This, however, is not the case for Instagram. Despite these limitations, the data still allows us to answer the posed research questions and draw conclusions about the investigated population.

The first research question aims at finding linguistic features that are stable across the two investigated platforms. As illustrated in Figure 1, both RE and sentence lengths are rather stable across both platforms which



one exception for each feature: Person D's RE score varies more than for the other investigated individuals, and Person C's sentence length differs between Twitter and Instagram much more than for the other three people. It thus appears as if syntactic complexity on social media platforms is worth investigating more closely for forensic linguistic purposes, much more so than the other features examined in this paper, which appear to be more variable.

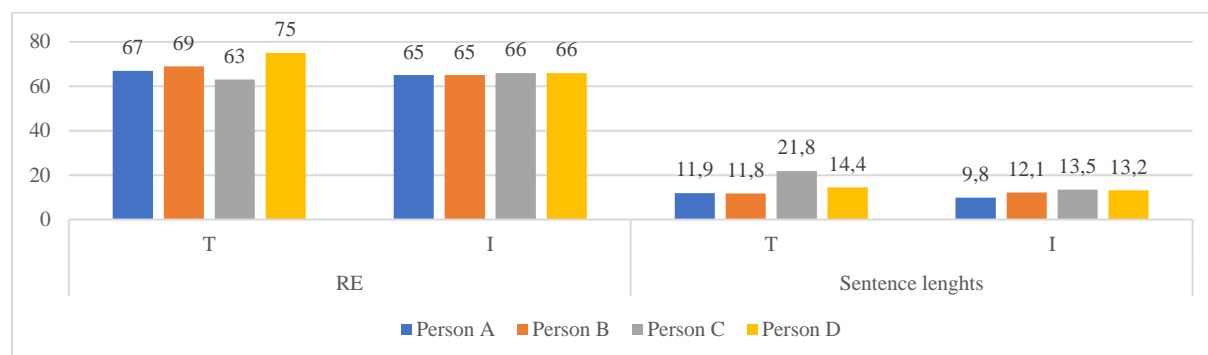


Figure 1. Comparisons of RE scores and sentence lengths (T=Twitter; I=Instagram)

In light of the search for platform-dependent features (research question 2), interesting findings are obtained. Figure 2 compares the use of punctuation and the presentation of numbers across all four individuals and both platforms. What can be seen is that both features vary between the platforms for all individuals, which makes these features unstable and of little use for authorship analysis. However, knowing that these features are dependent on platforms and likely to be variable for many people is still a helpful finding. Since the variation in the use of these features, however, is not consistent for the investigated individuals, meaning that no patterns typical for the platforms have emerged, no conclusions about the respective platform's genres can be drawn at this point.

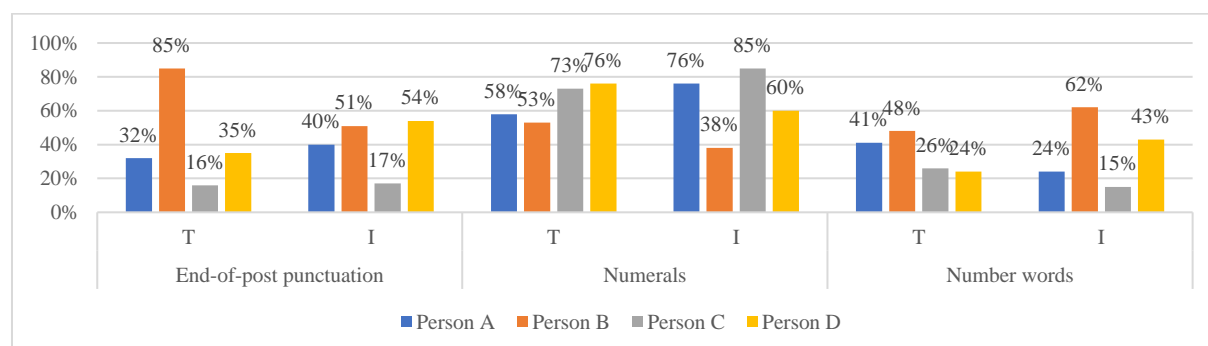


Figure 2. Comparisons of punctuation and number representation (T=Twitter, I=Instagram)

The discourse and CMC-features examined in this paper, however, are very promising in the light of our second research question and the attempt to detect platform-dependent features. Figure 3 shows an overview of the results for easier comparisons. The most important differences between the investigated individuals relate to the use of emoji, @mentions and hashtags. As shown in Figure 3, all three of these features are consistently more common on Instagram than on Twitter, which allows us to draw tentative conclusions about the genres of the examined platforms. That is, Instagram appears to invite users to use more CMC features, even though some of them, such as hashtags, have emerged on Twitter first (Zappavigna, 2013). Further, emoji seem to play a more important role in the messaging style on Instagram, as does the use of @mentions. The only feature for which no pattern was extracted is the use of capitalizations for emphasis. This, however, suggests that this feature needs to be investigated more closed in terms of individual, platform-independent, differences.

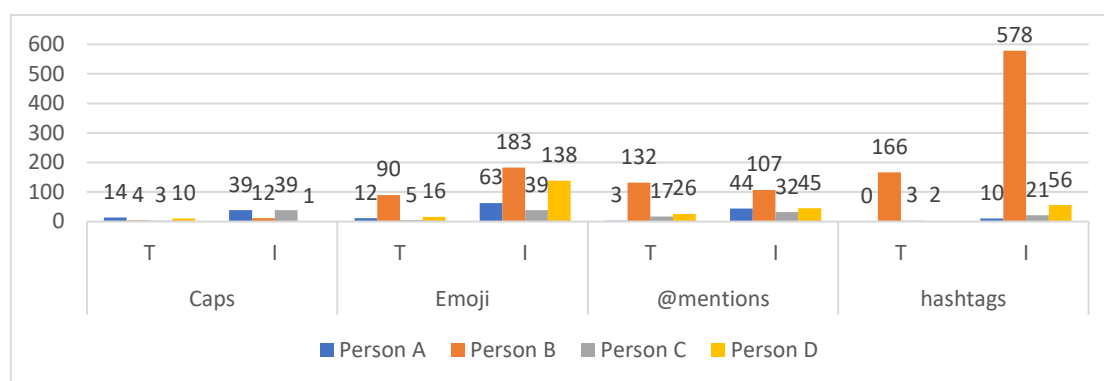


Figure 3. Comparison of discourse and CMC-features (T=Twitter; I=Instagram)

As Tannen (2013) has discussed, the medium by which an interaction occurs informs the metamessage. The metamessage as a concept can be traced to Bateson (1972:177-178), who states that “human verbal communication can operate and always does at many contrasting levels of abstraction”. Tannen’s interpretation of metamessages derives from Bateson’s idea of metacommunication. As it relates to CMC, Tannen illustrates that the choice of medium in communication between individuals, as in email vs. text messaging vs. Facebook, indicates certain communicative information, such as enthusiasm, terseness, or directness, among others. This ties in with the findings of the present study that shows differences in CMC-specific features related to the used platforms.

In line with this finding, examples (48) from Twitter and (49) from Instagram below show how the platform influences the form of the message in the examined dataset. The examples are taken from Person B:

(48) ““Happiness can only come from inside of you and is the result of your love. When you are aware that no one else can make you happy and that happiness is the result of your love, this becomes the greatest mastery of the Toltec: The Mastery of Love.” - @donMiguelRuiz”

(49) “Happiness can only come from inside of you and is the result of your love. When you are aware that no one else can make you happy and that happiness is the result of your love, this becomes the greatest mastery of the Toltec: The Mastery of Love.” - don Miguel Ruiz 🙏👉👈. This book is simple but powerful - thank you @aubreymarcus for the recommendation. At this point, most people know how much I love to read and share my books. Please share your favorites with me as well! 🙏 #bookclub #knowledge #bookworm#love #wordsofwisdom #toltec#toltecwisdom #donmiguelruiz”

As can be seen, the posts differ ever-so-slightly between the two platforms but convey the same semantic information. Furthermore, the examples illustrate neatly that the use of emoji and hashtags is higher on Instagram than it is on Twitter. This suggests that a person’s style is influenced by the used platform and that the platform itself becomes a medium that the user utilizes to achieve different interactional ends.

An important finding that complicates the use of emoji as markers of authorship (Marko, 2020) and that highlights platform-dependent usages relates to emoji functions on the platforms. Intriguingly, the use of emoji with a reinforcing function is much more common on Instagram than it is on Twitter for all investigated individuals, while the use of stand-alone emoji is overall more common on Twitter. These findings also suggest that emoji usage is heavily influenced by the platform, while at the same time preferences for specific emoji might remain stable – this, however, is an area that needs to be addressed by future research.

The findings presented here suggest that Twitter-style may be more analogous to SMS text messaging, while Instagram’s style might be more analogous to a blogging style (Zappavigna, 2016) in that Instagram’s metamessage conveys a more performative, identity-construction and documentation style than Twitter does. Whether these findings hold true for a larger population will have to be examined in future research.

## 6. CONCLUSIONS

This paper set out to investigate platform-dependent and platform-independent linguistic features for Twitter and Instagram to investigate both idiosyncratic linguistic features and platform-specific features. The findings obtained through the detailed quantitative and qualitative analysis of four individuals on both platforms reveal that while syntactic markers are rather promising stable features useful to forensic linguists, other features, such as CMC-specific ones, can inform researchers about the genres and metamessages of the investigated platforms. Capitalization for emphasis, or emphatic capitalization, is also likely to depend on individual differences, which requires more detailed research. These findings indicate that forensic linguists searching for idiolectal features are well advised to focus on morpho-syntactic features, while CMC-researchers investigating the genres and language use on social media platforms will profit more from focusing on features such as @mentions, hashtags, and emoji.

This is not to say that CMC-specific features are not subject to individual differences and therefore not at all useful to forensic linguists, but it does imply that those features are strongly variable and influenced by the used platform, which needs to be taken into account in an analysis. Many questions have been raised throughout the analysis and discussion of the results, which will be examined in future research. Nevertheless, both directions of research deserve further examinations with larger datasets to allow for more profound conclusions to be drawn.

## REFERENCES

1. Ayers, J., Caputi, T., Nebeker, C. & Dredze, M. (2018). Don't quote me: Reverse identification of research participants in social media studies. *Digital Medicine* 1(1), 1-2. article 30. DOI:10.1038/s41746-018-0036-2.
2. Bateson, G. (1972). A theory of play and fantasy. In G. Bateson (ed.) *Steps to an Ecology of Mind*. New York: Ballantine, pp.177-93.
3. Bell, A. (1984). Language style as audience design. *Language and Society* 13(2), 145-204.
4. Biber, D., & Egbert, J. (2018). *Register Variation Online*. Cambridge: CUP.
5. boyd, d. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. In Z. Papacharissi (ed.) *Networked Self: Identity, Community, and Culture on Social Network Sites*. London: Routledge, pp.39-58.
6. boyd, d., Golder, S. & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *Proceedings of the 43rd Annual Hawaii International Conference on System Sciences* 1-10. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5428313](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5428313)
7. Brownlow, S., Pate, M., Alger, A. & Naturile, N. (2019). What were they thinking? Analytic and cognitive language in Instagram captions. *Social Behavior Research and Practice* 4(1), 21-25. DOI: 10.17140/SBRPOJ-4-117
8. Bruns, A. & Moe, H. (2014). Structural layers of communication on Twitter. In K. Weller, A. Bruns, J. Burgess, M. Mahrt & C. Puschmann (eds), *Twitter and Society*. New York: Peter Lang, pp.15-28
9. Chaski, C. (2001). Empirical evaluations of language-based author identification techniques. *Forensic Linguistics* 8(1), 1-65.
10. Coulthard, M., Johnson, A. & Wright, D. (2017). *Introduction to Forensic Linguistics*. London, England & New York, NY: Routledge.
11. Danesi, M. (2016). *The Semiotics of Emoji*. E-book. Bloomsbury.
12. Dayter, D. (2015). Small stories and extended narratives on Twitter. *Discourse, Context and Media* 10, 19-26.
13. Ehrhardt, S. (2018). Authorship attribution analysis. In J. Viscanti & M. Rathert (eds), *Handbook of Communication in the Legal Sphere*. Berlin, Germany: deGruyter, pp.169-200.
14. Evans, V. (2017). *The Emoji Code*. London: Picador.
15. Fairclough, N. (1992). *Discourse and Social Change*. Cambridge: Polity Press.
16. Fairclough, N. (1995). *Media Discourse*. London: Edward Arnold.
17. Flesch, R. (2016). *How to write plain English*. University of Canterbury. Archived from the original on July 12, 2016. Retrieved February 5, 2016.
18. Gawne, L. & McCulloch, G. (2019). Emoji as digital gestures. *Language@Internet*, 17.
19. Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
20. Grant, T. (2010). Text messaging forensics. TxT 4n6: Idiolect free authorship analysis? In M. Coulthard & A. Johnson (eds), *The Routledge Handbook of Forensic Linguistics*. London: Routledge, pp.508-522
21. Grant, T. (2013). TxT 4n6: Method, consistency, and distinctiveness in the analysis of SMS text messages. *Journal of Law and Policy* 21(2), 467-494.
22. Grant, T. & MacLeod, N. (2020). *Language and Online Identities. The Undercover Policing of Internet Sexual Crime*. Cambridge: CUP.
23. Hardaker, C. & McGlashan, M. (2016), "Real men don't hate women": Twitter rape threats and group identity. *Journal of Pragmatics* 91, 80-93.
24. Herring, S. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In S.A. Barab, R. Kling & J.H. Gray (eds.), *Designing for virtual communities in the service of learning*. New York: CUP, pp.338-376.
25. Herring, S. & Stoerger, S. (2013). Gender and (a)nonymity in computer-mediated communication. In J. Holmes J, M. Meyerhoff & S. Ehrlich (eds.), *Handbook of Language and Gender*. Hoboken: Wiley-Blackwell Publishing, pp.567-586.
26. Ishihara, S. (2017). Strength of linguistic text evidence: A fused forensic text comparison system. *Forensic Science International* 278, 148-197.
27. Jenkins, H., Ford, S. & Green, J. (2013). *Spreadable Media: Creating Value and Meaning in a Networked Culture*. New York: New York University Press.
28. Kemp, S. (2019). Digital 2019: Global internet use accelerates. Accessible at <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>

29. Krieg-Holz, U. & Bülow, L. (2016). *Linguistische Stil- und Textanalyse*. Tübingen: narr.
30. Layton, R., Watters, P. & Dazeley, R. (2010). Authorship attribution for Twitter in 140 characters or less. *IEEEI*, 1-8.
31. Leaver, T., Highfield, T. & Abidin, C. (2020). *Instagram: Visual Social Media Cultures*. Cambridge, England: politybooks.
32. MacLeod, N. & Grant, T. (2012). Whose tweet? Authorship analysis of micro-blogs and other short-form messages. *Proceedings of the International Association of Forensic Linguists' 10th Biennial Conference*. Aston University, Birmingham, UK, 210-224.
33. Marko, K. (2020). Exploring the distinctiveness of emoji use for digital authorship analysis. *Language and Law / Linguagem e Direito*, 7(1-2), 36-55.
34. Marwick, A. (2013). Online Identity. In J. Hartley, J. Burgess & A. Bruns (eds), *Companion to New Media Dynamics*. Blackwell Companions to Cultural Studies. Malden: Blackwell, pp.355-365
35. Marwick, A. & boyd, d. (2010). I tweet honestly, I tweet passionately: Twitter users, content collapse, and the imagined audience. *New Media & Society* 13(1), 114-133.
36. McCulloch, G. (2019). *Because Internet*. New York: Riverhead Books.
37. McMenamin, G. (2002). Forensic Linguistics. *Advances in Forensic Stylistics*. Boca Raton: CRC Press.
38. McMenamin, G. (2010). Forensic stylistics. Theory and practice of forensic stylistics. In M. Coulthard & A. Johnson A (eds), *The Routledge Handbook of Forensic Linguistics*. London: Routledge, pp.487-507
39. Nguyen, D., Gravel, R., Trieschnigg, D. & Meder, T. (2013). "How old do you think I am?": A study of language and age in Twitter. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, 1-10.
40. Orebaugh, A. & Allnutt, J. (2009). Classification of instant messaging communications for forensic analysis. *The International Journal of Forensic Computer Science* 1, 22-28.
41. Paige, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication* 6(2), 181-201.
42. Paige, R., Barton, D., Unger, J. & Zappavigna, M. (2014). *Researching Language and Social Media*. London and New York: Routledge.
43. Peersman, C., Daelemans, W. & van Vaerenberg, L. (2011). Predicting age and gender in online social networks. *SMUC* 37-44.
44. Schneebeli, C. (2017). The interplay of emoji, emoticons, and verbal modalities in CMC: a case study of YouTube comments. *VINM 2017 Visualizing (in) the Social Media*, 1-15.
45. Sousa Silva, R., Laboreiro, G., Sarmiento, L., Grant, T., Oliveira, E. & Maia, B. (2011). 'Twazn me!!!'; ('Automatic authorship analysis of micro-blogging messages. In Munoz R, Monotony A, Métais E (eds) *NLDB*. Berlin: Springer, pp.161-168
46. Tagg, C. (2015). *Exploring Digital Communication*. Oxon: Routledge.
47. Tannen, D. (2013). The medium is the metamessage: Conversational style in new media interaction. In D. Tannen & A. M. Trester (eds), *Discourse 2.0: Language and New Media*. Washington, DC: Georgetown University Press, pp.99-118
48. Townsend, L. & Wallace, C. (n.d.). *Social media research: a guide to ethics*. University of Aberdeen, 1-16.
49. Van Dijck, J. & Poell, T. (2013). Understanding social media logic. *Media and Communication* 1(1), 2-14. DOI: 10.12924/mac2013.01010002
50. Veum, A. & Undrum, L.V.M. (2018). The selfie as a global discourse. *Discourse & Society* 29(1), 86-103. DOI: 10.1177/0957926517725979.
51. Zappavigna, M. (2013). *Discourse of Twitter and Social Media*. London, England & New York: Bloomsbury.
52. Zappavigna, M. (2016). Social media photography: construing subjectivity in Instagram images. *Visual Communication* 15(3), 271-292.