Master's thesis

Language on Twitter. An analysis of (in) formality of tweets: Study on Carlsberg, LAY's and Toblerone (FMCG industry)

Radboud Universiteit Nijmegen

Miruna Stefancu

08-10-2015

Abstract

This study aimed to shed light on Twitter from a corporate communication and a linguistic perspective. The main focus was on the 'message style framework' (Cornelissen, 2008), a traditional tool used in strategic communication campaigns, which consists of 5 message styles – rational, emotional, symbolic association, generic and preemptive, but an extra message style – 'No message style' – was added in order to code tweets that did not fit into any of the style categories. The message style framework was correlated with the (in) formality level into formal (above 50%) or informal (below 50%) categories, based on Heylighen and Dewaele's (2002) *F-score* (below 50% = informal; above 50% = formal). The framework was also correlated with the occurrence of atypical use of special characters, punctuation or capitalisation, also known as paralinguistic language. The framework applicability on Twitter was analysed in a corpus of 120 tweets, posted by 3 FMCG industry brands: Carlsberg, LAY's and Toblerone. The relationship between the brand and the style category was tested as well. Results indicated a substantial number of tweets were placed under no message style, followed by the emotional message style. Most of the tweets were formal, but no significant relationship between the (in) formality level and the message styles was found. There was a high occurrence of paralinguistic language, but no significant relationship between the (in) formality level of the message styles and the paralinguistic language. The relationship between the message style and the brand was significant. Carlsberg posted the most tweets from almost all the message styles. The study indicated companies are yet to implement the message style framework in their online communication strategy and efficiently correlate it with the (in) formality level and paralinguistic language in order to make their online communication campaigns as successful as the traditional ones.

Introduction

In the last 10 years, Internet has taken over the traditional means of corporate communication. The traditional communication channels − television, radio and newspapers – have been rapidly replaced by social media platforms such as Facebook, Twitter, LinkedIn, and YouTube. The greatest advantage of these platforms is the easy access to information in short time and the reach to a wider audience (followers), given the global availability of Internet.

A new chapter has started for corporate communication with regard to the use of social media as part of communication strategy and planning. Corporate brands engage in an online activity to spread information easily across existing stakeholders and can also attract new ones. This activity is referred to as Computer Mediated Communication (CMC). By definition, CMC is "communication that takes place between human beings via the instrumentality of computers" (Thurlow, Lengel & Tomic, 2004, p.83). The flexibility and convenience of CMC make it often useful to coordinate different communication situations (Norris, 2012). The present study aimed to determine whether the functionality of the 'message style framework' (Cornelissen, 2008), the (in)formality level and the occurrence of paralinguistic language differs between traditional, offline communication and communication on Twitter as a form of CMC. Twitter is a communication channel that has increased in popularity, leading to studies such as Jamal and Waters' (2011), Messner et al.'s (2013) and Stelzner's (2009), who all indicated that Twitter is the most used social media channel in strategic communication campaigns. Other studies, such as Lovejoy et al.'s (2010) and Muralidharan, Rasmussen, Patterson and Shin's (2011) indicated the use of Twitter fails to build communities, despite the channel's popularity and growing importance in communication campaigns.

The current research did not aim to prove or deny Twitter's ability to build communities, but to emphasise the use of the channel to engage followers through the message style framework and the framework's correlation with the (in) formality and paralinguistic language of tweets.

Message style framework

The corporate communication 'message style framework' (Cornelissen, 2008) is a key element of the communication strategy, which lies at the heart of the corporate communication.

The communication strategy was originally used in offline mediums and describes the general image that an organisation aims to project through themed messaging to stakeholders (Cornelissen, 2008). Themed messages relate to particular 'themes', such as capabilities, strengths or values of an organisation. These themes are further classified in the following five styles: rational message style, symbolic association message style, emotional message style, preemptive message style, and generic message style (Cornelissen, 2008).

To begin with, the *rational* message style is generally used when an organisation wants to make a superiority claim about its products or achievements based upon a distinctive advantage in its capabilities, size or resources. The message style follows a basic argumentation structure where the grounds for the claim for superiority are delivered through supporting information. Zhang, Sun, Liu and Knight (2014, p. 2106) illustrate the example of a Starbucks advertisement that emphasises a sandwich's healthy ingredients, which suggests Starbucks is superior to its competitors by using healthy ingredients in the preparation of sandwiches (distinctive advantage).

Next, the *symbolic association* message style is based on psychological differentiation through symbolic association. More precisely, the organisation aims to identify a set of symbols and values and later associate them with the image of the organisation. Examples of this message style are found in the way organisations associate themselves with a sport or a good cause in sponsorship activities and in explicitly stating values and attributes that serve as guidance in the organisation's conduct. Gordon (2002, p. 431) cites the example of Coca-Cola, which, starting from the 1900s, has included healthy and lively young women in their advertising campaigns, suggesting Coca-Cola is associated with a healthy and active lifestyle.

Another type of symbolically oriented communication is the *emotional* message style. This style attempts to engage an organisation's stakeholders by evoking reactions through different emotions including romance, nostalgia, excitement, joy, fear, guilt, disgust or regret. For instance, Moore et al. (1995) cite the example of a Hallmark greeting card advertisement, which amplifies a person's most important moments in life, indicating the emotions that are stirred through the advertisement, such as joy, sadness, excitement or romance.

Following, an organisation can also use a *generic* message style in its communication. This style is preferred by organisations that are dominant in an industry and want to make a straight claim without declaring superiority. The generic message is most suitable

when the intention is to make the brand synonymous with the product category (Clow & Baack, 2007, p.199). To support this, Clow and Baack (2007, p.199) cite the example of Nintendo, which is the leader of the handheld game category, with a market share of more than 98%. That is why, the authors say, when consumers think about the handheld game industry, the first brand that comes into their minds is Nintendo.

Lastly, organisations can use a *preemptive* message style. Like the generic message style, the preemptive style is used to make a generic claim, but with suggestion of superiority. It is particularly useful when a company wants to prevent its competitors from saying the same. For instance, Clow & Baack (2007, p. 200), cite the example of the Crest toothpaste, which is known as the 'cavity fighter'. Because of the renowned association of the Crest brand with this attribute, other companies cannot make the same statement, even though all toothpastes have the same characteristic.

In the current study, the message style framework was applied on Twitter to check whether tweets can be classified in the style categories in the same way styles apply in traditional, offline corporate communication.

Formality and informality

Heylighen and Dewaele (1999) indicate a universal definition of formality does not exist, but cite the definition of the *formal speech* from the Dictionary of Language Teaching and Applied Linguistics: "the type of speech used in situations when the speaker is very careful about pronunciation and choice of words and sentence structure. This type of speech may be used, for example, at official functions, and in debates and ceremonies" (Richards, Platt & Platt, 1997, p.144).

In a later study of Heylighen and Dewaele (2002), the concepts of *formality* and *informality* are based on the context of communication situations. The authors describe *formal language* as structurally more complex than the informal language ─ which they call *contextual speech* ─ because "formal expressions require more time, attention and cognitive processing to be produced and understood" (Heylighen & Dewaele, 2002, p. 302). Compared to the formal language, contextual speech contains less, shorter and more frequent words, which are easily retrieved, since the context shared by the sender and receiver provides the additional information lacking in the linguistic expression itself. Also, contextual speech is "more interactive or involved, reaching directly to interlocutors […]" (Heylighen & Dewaele 2002, p. 302). Irvine (1979) argues that in (formality) can also describe the characteristics of a

social situation, and not necessarily the kind of code used in that situation. The relevant characteristics of the situation may have something to do with the vocal tone, so that a formal situation requires a display of seriousness, politeness and respect. Heylighen and Dewaele (2002) introduced a method to evaluate the formality and informality of a message, called the *F-score*. The frequency of the non-deictic category of words (nouns, adjectives, prepositions, and articles) is likely to increase with the formality of a text, while the frequency of the deictic word category (pronouns, verbs, adverbs, and interjections) is expected to increase with the contextuality of a discourse (Heylighen & Dewaele, 2002, p. 309).

Language in traditional, written communication is generally formal, as opposed to the oral communication, which is always informal (Bortolini et al., 1971; Heylighen & Dewaele, 2002; Hudson, 1994; Juilland & Traversa, 1973). Heylighen and Dewaele (2002), based on the frequency dictionaries of Bortolini et al. (1971), Juilland and Traversa (1973), and Hudson (1994) indicate how for instance movies, phone conversations, spontaneous speeches and conversations are informal compared to essays, technical and scientific articles, and informational writing, whose *F-scores* indicate a higher level of formality. For example, on Heylighen and Dewaele's (2002) scale, technical and scientific texts scored 71.6% (p. 314) and on Hudson's (1994) scale informational writing had an F-score of 61% (more formal than any form of spoken communication).

In relation to the message style framework (Cornelissen, 2008), it was already expected that each of the component message style would possess a different level of formality, which can vary across examples of communication. To measure the (in) formality level of such instances, an adaptation of Heylighen and Dewaele's (2002) *F-score* was used (formal = above 50%; informal = below 50%).

Paralinguistic language

Schuller et al. (2013) define paralinguistics as the "discipline dealing with those phenomena that are modulated onto or embedded into the verbal message, be this in acoustics (vocal, non-verbal phenomena) or in linguistics (connotations of single units or of bunches of units)" (p. 5). The authors explain how laughter can express a certain state of mind, an emotion or a mood, how different denotations of a word can reveal the social class and/or the character, and how the use of adjectives can point toward personality traits or emotional states (Schuller et al., 2013, p.5). Tench (1990) and Bombelli et al. (2013) also explain that paralinguistic features are expressed through intonation and are present in the interpersonal communication,

and reflect the emotional state of the interlocutors. Brown (1990) defines paralinguistic features as characteristics of speech that "contribute to the expression of attitude by a speaker and do not form an intrinsic part of the phonological contrasts which make up the verbal message" (p.112). In the present study, which is based on Twitter, the traditional paralinguistic features are replaced by strategies such as "reduced or simplified speech; slow, exaggerated pronunciation and intonation; short sentences; special lexicons […]" (Murray, 2000, p. 401), use of abbreviations, typographical and spelling errors or emoticons (Ferrara, Brunner, & Whittemore, 1991; Gains, 1999; D. E. Murray, 1991).

Several linguistic studies on Twitter and tweets have been conducted before; however, none of them focused on the applicability of the message style framework on Twitter and the level of formality of each component message style. Also, none of these studies investigated a possible link between the (in) formality level of the message styles and the use of paralinguistic language, which is exactly the aim of the current study. Considering the character-constraint of Twitter, differences from the traditional, offline language were prefigured. Yates' (1996) study on CMC indicated that the nature of CMC is mostly similar to the spoken language, but contains some features from the written language as well, such as lexical density. Lexical density is "a measure of the proportion of lexical items (i.e. nouns, verbs, adjectives and some adverbs) in a text" (Johansson, 2008, p.61). Similarly, Yates and Orlikowski (1993) found that CMC language is comparable to both the oral communication (hence informal) and the written communication. Mosquera and Moreda (2011) also stated that current online language is informal, being characterised by "non-standard abbreviations, colloquial expressions or presence of slang words" (p. 186). Therefore, it would be expected language on Twitter, which is a form of CMC, is relatively more informal than written communication. Walther and Burgoon (1992) found that although written messages may be more formal than oral communication, informality might increase as "users are likely to develop and imbue their messages with informality cues as they become accustomed to each other and the medium" (p. 60 and p. 76). The present study aimed to investigate whether tweets have similar characteristics as determined in the past studies or the linguistic register (formal/informal) is fundamentally different.

Paralinguistic language has existed long before the introduction of computers and CMC. Pak and Paroubek (2010), Schandorf (2013) and Smith (2013) found the presence of paralinguistic language in tweets is high. Schandorf (2013) stresses paralinguistic language is typical for the fast and telegraphic new media communication. Pak and Paroubek (2010) found Twitter users seem to display a lot of emotion into the messages they send, by

including paralinguistic language features, such as happy or sad emoticons. Since Twitter is a text-based social media, text styles, for instance all caps and punctuation (especially exclamation signs) might play the role of substituents of vocal expressive gestures (Schandorf, 2013). Those text styles, including expressive typography, make up the paralinguistic language, as also described by Smith (2003).

Fast Moving Consumer Goods (FMCG)

To tackle the goal of the study, tweets from the Fast Moving Consumer Goods (FMCG) industry – 3 different brands: *Carlsberg*, *LAY's* and *Toblerone* – were chosen for analysis. The FMCG industry was the most suitable choice because FMCG is a sector where brands can live forever. For instance, Procter & Gamble management believes that within the consumer area, if brands are well managed, they should last forever (Moss & Schuiling, 2004). The FMCG industry is defined as "relatively inexpensive, frequently purchased and rapidly consumed items on which buyers exert only minimal purchasing effort" (Dibb, Simkin, Pride & Ferrell, 2006, p. 298). Moreover, as Dovey (2009) argues, customers must become involved with FMCG brands and brands "must follow their customers online however hard it is" (p. 15). The impact of traditional communication channels is lessening as customers spend more time online. They are becoming more involved with FMCG brands and even have the power to influence them, as it was seen in numerous campaigns initiated by FMCG brands, which are known to have always put emphasis on understanding their customers (Dovey, 2009). *Carlsberg*, *LAY's* and *Toblerone* are representative brands for the FMCG industry because they are well known in the offline and online medium. The informal nature of the FMCG led to the expectation of a high level of informality and presence of paralinguistic language and to the use of a variation of message styles.

So far, the message style framework has only been applied by companies using traditional communication channels. The current investigation's aim was to determine if Twitter is a suitable medium for organisations from the FMCG industry to integrate the message style framework in their corporate communication strategy and whether these organisations can adapt the message styles to a specific linguistic register (formal/informal), and paralinguistic features that fit with the corporate image or goals. Bearing this in mind, the following research questions were formulated:

RQ1. Is there evidence Cornelissen's (2008) message style framework can also be applied to tweets from organisations from the FMCG?

RQ1 a. What is the relationship between the message style categories and the FMCG brands?

RQ2. What is the level of formality of each message style within the framework?

RQ3. What is the relationship between the level of formality of each message style within the framework and the paralinguistic features?

Method

To investigate how formality and informality on social media links with the message style framework of corporate communication, and how paralinguistic features relate to formality and informality, a *corpus analysis* of tweets was done. Since the present study focused on the interpretation of written content (tweets), the procedure was based on Herring's methodological approach to computer-mediated discourse, called *computer-mediated discourse analysis (CMDA)*. The language-focused content analysis can be either *qualitative* (observations of discourse phenomena in a sample of text) or *quantitative*, where phenomena may be coded and counted and summaries of their relative frequencies calculated. Sometimes, the quantitative approach may contain a qualitative component; for instance, "in deciding what counts as an instance of a phenomenon to be coded and counted […]" (Herring, 2004, p. 343). In the present study, the sample was first coded and counted, and further analysed from a linguistic perspective, which is why the corpus analysis was both quantitative and qualitative.

Materials

The three Twitter accounts chosen for further investigation belong to three famous brands from the FMCG industry: Carlsberg, LAY's and Toblerone. The brands accounts were chosen on 5 criteria: *originality, verification, internationality, activeness* on Twitter and *similarity of industry*. Only *original* tweets posted by the companies were taken into consideration; no retweets by either of the three FMCG brands or their followers were included. The tweets to be studied were selected from the period December 2014 – April 2015.

As mentioned, only *verified* accounts were chosen, as a measure of safety and authenticity for followers (consumers). Verified accounts can be easily identified by the blue verified badge on their Twitter profile. In order to be able to extrapolate the results of the research, the accounts were chosen from one industry only. The accounts were also chosen based on the *internationality* of the brands. All three FMCG industry brands are well known

and reach out to consumers from all over the world. Lastly, to make the results comparable, the activeness of the three accounts had to be similar. Features such as number of followers and number of tweets indicate how large the audience of the accounts is, how many followers are exposed to the tweets, and how popular the brand is amongst followers/consumers. All three FMCG brands had a similar number of followers and a similar number of tweets, which made them suitable for the investigation.

Procedure

The corpus analysis commenced with the selection of 40 tweets from each FMCG brand, which made a total data of 120 tweets. To answer to the first research question – whether the message style framework can also be applied to tweets from organisations from the FMCG – tweets were coded and grouped according to Cornelissen's (2008) message style framework, to determine the *general sample*. Tweets were coded and grouped in order to establish what qualifies tweets to one of the five message styles – rational, symbolic association, emotional, generic or preemptive. To simplify the coding, the symbolic association message style was renamed 'symbolic'. However, it was expected that a part of the sample would not fit into any of the five message styles or it would be a combination of message styles. To avoid any confusion, a *no message style* category was added. The coding revealed no tweet fit into the rational message style category, which is why no example of a rational tweet will be provided, only illustrations of the other message styles (No message style included).

In the following paragraphs, an overview of each message style will be provided, with a detailed explanation of what characterises a tweet in a certain message style. In all the cases, the classification was based on language structures or characteristics that indicated the affiliation of the tweet with the message style. More illustrations of tweets are provided in Table 7, in the Appendix 1.

Tweet (15): preemptive style

Carlsberg: *Probably the best ice bottle of @carlsberg in the world. Snow Dome in Levi Finland*

This tweet was classified as preemptive due to the association between the Snow Dome in Levi Finland and the "best ice bottle of Carlsberg in the world", which is a generic illustration, but with suggestion of superiority. This way, the beer company is trying to hinder its competitors from making the same claim. The integration of Carlsberg's slogan ("Probably

the best beer in the world") is an additional contribution to the preemptiveness of the message.

Tweet (40): generic style

Carlsberg: 37 reasons why Denmark is awesome. Probably hard to argue with 29! No. 2 is Stockholm though, @BuzzFeed ;) http://bzfd.it/1D6ZwrS

In another example of Carlsberg, although there is no direct relationship suggested between the beer and the "awesomeness", it can be implied that the beer brand – which claims to "probably be the best in. the world", thus dominant in the industry – aims at making a straight claim without implying superiority. Based on this type of claim, the tweet was regarded as generic.

Tweet (3): emotional style

Carlsberg: Incredibly romantic or extremely eccentric? Read our best beer stories, follow us on Instagram @carlsbergsince1847. http://bit.ly/1wghPL1

In this tweet, Carlsberg is trying to involve its followers in pursuing the company's online activity on Instagram (a popular social network on which users post photographs accompanied by hashtags, a similar principle to Twitter's), by provoking certain reactions through emotions; as it can be seen, the tweet is addressed to followers who are "incredibly romantic" or "extremely eccentric", which inscribes the tweet into the emotional message style category. Furthermore, the tweet is an evident illustration of dialogue with the followers, because the message commences with a rhetorical question ("Incredibly romantic or extremely eccentric?"), followed by action verbs ("read", "follow").

Tweet (20): no message style

Carlsberg: Maybe right here @gilchristdm. #Probably http://gph.is/1LdCeW5

Since this tweet was a direct reply to a follower (@), there was no relevant information contained that would place the tweet under a category of message style.

Tweet (46): emotional style

LAY's: Holiday travel can be torture. We've all been there. Reward yourself with your favorite cocktail and snack.

Like Carlsberg's tweet inviting "incredibly romantic or extremely eccentric"

followers to engage with the organisation's other online activity, this tweet posted by LAY's was placed in the emotional message style as well. The organisation identifies itself with the followers ("We've all been there") when making reference to the holiday travel as being "a torture". Moreover, LAY's is inviting its followers to become more engaged with the brand by overcoming the bad emotions (the "torture"). Consuming the crisps ("favorite […] snack") might help the followers replace the bad emotions with positive ones ("Reward…").

Tweet (49): no message style

LAY's: #DoUsAFlavor is back! Submit Lay's flavors NOW at http://dousaflavor.com for a chance to win $1 million! See rules.

This particular tweet posted by LAY's does not seem to fit in any message style due to the objectivity of the message; as most of LAY's tweets that were analysed, the information transmitted by this tweet is about a contest organised by the company on a period of several months.

Tweet (105): emotional style

Toblerone: Sources close to us claim Toblerone has been spotted at the airport. We'll keep you posted… #DeadlineDay

Even deprived of visual representations, such as photos or videos as addition to the text message, this particular tweet consists of a metaphor ("Toblerone has been spotted at the airport") and it aims at making followers engaged with the Twitter activity of the chocolate brand ("We'll keep you posted…"). Since the tweet is meant to evoke suspense emotions among followers, it was classified as emotional.

Tweet (91): no message style

Toblerone: Three is the magic number.

Without any additional information such as photos or videos, this tweet seems taken out of its context and can be of any significance to the followers, hence the difficulty of categorising it in a message style category.

To answer to the second research question, the level of formality of the general sample of tweets was determined using the *F-score* (Heylighen & Dewaele, 2002). As mentioned in the literature review, nouns, adjectives, prepositions, and articles are parts of the

speech which are likely to increase the formality of language, while pronouns, verbs, adverbs, and interjections are indicators of informality.

First of all, every tweet was checked for parts of speech using Parts-of-speech.Info (http://parts-of-speech.info/#{"page":"tagging"}). To rule out any form of possible inaccuracy, the tweets were double-checked manually. Subsequently, the formula was calculated to determine the formality scale for each tweet. The *F-score* can be calculated using the following formula:

$$F = (\text{noun frequency} + \text{adjective freq.} + \text{preposition freq.} + \text{article freq.} - \text{pronoun freq.} - \text{verb freq.} - \text{adverb freq.} - \text{interjection freq.} + 100)/2$$

The frequencies are expressed by percentages of the number of words belonging to a particular category with reference to the total number of words in that sequence. The value of F will always vary between 0% and 100%, but will never reach these limits. If the formality of a language excerpt is high, the value of F will increase. For example, Heylighen and Dewaele (2002) conducted a study in which they collected a corpus of two speech styles and one written style (all containing approximately 30,000 words), in three situations in decreasing order of formality: an informal conversation; an oral examination testing the knowledge of a language, and an essay. The results indicated the following values of the *F-scores*: 44% (conversation), 54% (examination) and 56% (essay). In accordance with the authors' expectations, the frequency of nouns, adjectives, articles, and prepositions increased with an increase of formality, whereas the frequency of pronouns, adverbs, and verbs decreased (Heylighen & Dewaele, 2002). In the present study however, due to coding reasons, the decision was made to treat scores *under 50%* as *informal* and scores *higher than 50%* as *formal*; nonetheless, in the event of obtaining the exact 50% value, the formality was established based on the frequencies of the parts of speech contained in the tweet. For example, if the result of the *F-score* was 50%, the tweet was classified as informal if the frequencies of the pronouns, verbs, adverbs, and interjections were higher than those of the nouns, adjectives, prepositions, and articles. In the event of the opposite situation, tweets were classified as formal.

Lastly, to answer to the third research question, the correlation between the level of formality of the general sample and the presence of paralinguistic features was determined. Word capitalisation, emoticons, onomatopoeia (interjections), punctuation (extensive use of punctuation to indicate pitch and intonation or express surprise), substitution – use of abbreviated forms of a word (*u* = you, *2* = too/to, *ic* = I see, *y* = yes, *r.u.* = are you; *info*, *tech*)

(Smith, 2003, p. 46; Ferrara, Brunner, & Whittemore, 1991; Gains, 1999; D. E. Murray, 1991), acronyms (*IMHO* = in my humble opinion; *F2F* = face-to-face) or typographical and spelling errors (Ferrara, Brunner, & Whittemore, 1991; Gains, 1999; D. E. Murray, 1991) were coded as paralinguistic language features.

In order to establish the degree of inter-reliability of the collected data, Cohen's kappa was calculated by the researcher and by an independent coder with the statistical programme SPSS 21. Both coders had to analyse at least 10% of the total sample, namely a minimum of 12 tweets (10% of 120 = 12 tweets). In total, seven Cohen's kappa were calculated – five for the each message style (including the 'no message style' category), one for the formality level and one for the paralinguistic language features.

Firstly, several Cohen's kappa were run to determine whether there was an agreement between the two coders about the categorisation of the tweets into each message style. The values obtained indicated that the interrater reliability of the variables 'symbolic', 'preemptive', and 'no message style' were satisfactory: $\kappa = .73$, $p < .001$, $\kappa = .76$, $p < .001$ and $\kappa = .71$, $p < .001$, respectively. The variables 'emotional' and 'generic' had a good interrater reliability: $\kappa = .80$, $p < .001$ and $\kappa = .85$, $p < .001$. As none of the tweets were rational, no Cohen's kappa was calculated for the variable 'rational'. Secondly, another Cohen's kappa was run to check the agreement for the variable 'formality'. There was perfect agreement between the coders, $\kappa = 1.00$, $p = .001$. Lastly, the interrater reliability of the variable 'paralinguistic language features' was substantial (Landis & Koch, 1977): $\kappa = .62$, $p = .020$.

Statistical treatment

Several types of descriptive statistics analyses were performed. Frequencies were run in order to determine the proportions of message styles and of formality. Chi-square tests were run so as to establish whether there was a statistical relationship between the message styles and the level of formality and between the paralinguistic language features and the (in) formality level.

Results

The first research question involved the applicability of the message style framework on tweets belonging to organisations from the FMCG industry. To determine this, a descriptive

statistics frequencies test was run on the message styles (no message style category included). The results can be seen in Table 1.

Table 1: Frequency scores and means of the message styles (*N* = 120).

	Count	%
Rational	0	0
Symbolic	13	10.8
Emotional	37	30.8
Generic	3	2.5
Preemptive	**7**	5.8
No message style	60	50

Additionally, a Chi-square test was carried out in order to determine whether there was a significant relationship between the variables 'company' and 'message style'. The test revealed that there was a significant relationship between the company and the message styles – no message style category included (χ^2 (8) = 29.90, $p < .001$). Toblerone posted the highest number of symbolic tweets, while LAY's posted the lowest. The most frequent emotional tweets were posted by Carlsberg, whereas the least were posted by LAY's. Carlsberg also posted the most generic tweets; however, no generic tweets were found on LAY's account. Similarly, the most preemptive tweets belonged to Carlsberg, but none of them were posted by LAY's. In contrast, LAY's posted the most tweets of no message style, whereas Carlsberg posted the least of them. A Fischer's Exact Test indicated a value of 28.60, $p < .05$. Further results are available in Table 2.

Table 2: Company * Message style Crosstabulation

			Message style				
			Symbolic	Emotional	Generic	Preemptive	No message style
Company	Carlsberg	Count	4	19	2	6	9
		% within Company	10.0%	47.5%	5.0%	15.0%	22.5%
		% within message_style	30.8%	51.4%	66.7%	85.7%	15.0%
	LAY'S	Count	3	6	0	1	30
		% within Company	7.5%	15.0%	0.0%	2.5%	75.0%
		% within message_style	23.1%	16.2%	0.0%	14.3%	50.0%
	Toblerone	Count	6	12	1	0	21
		% within Company	15.0%	30.0%	2.5%	0.0%	52.5%
		% within message_style	46.2%	32.4%	33.3%	0.0%	35.0%

An additional step was performed to assist in answering the first research question. The average length of tweets was calculated in Microsoft Excel, to get an overview of the impact of the character constraint of Twitter compared to the traditional, offline communication messages, where there is more freedom and no character limit. The values can be found in Table 3.

Table 3: Length of tweets (average number of words per tweet) (*N* = 120)

Length of tweets	*M*	*SD*
	5.82	14.1

In order to answer to the second research question of the present study, which concerned the level of formality of the style categories, several steps were followed. Firstly, once the tweets were grouped according to the message style they fit in, their level of formality was calculated manually using the *F-score* formula. Subsequently, in a similar way to the first research question, a descriptive statistics frequencies test was carried out so as to determine the proportions of formal and informal tweets respectively. The test indicated that

31 (25.8%) tweets were informal and 89 (74.2%) tweets were formal (*M* = .74, *SD* = .44). Furthermore, a Chi-square test was carried out to conclude whether there was a significant relationship between the variables 'level of formality' and the message styles – no message style category included. The test showed that there was no significant relationship between the level of formality (formal/ informal) and the message styles − no message style category included ($\chi 2$ (4) = 2.93, *p* = .590). A Fisher's Exact Test indicated a value of 3.06, *p* = .554. The crosstabulation below provides further results (Table 4).

Table 4: Level of formality * Message styles Crosstabulation

			Message styles				
			Symbolic	Emotional	Generic	Preemptive	No message style
Level of formality	Informal	Count	4	9	1	0	17
		% within Formal_informal	12.9%	29.0%	3.2%	0.0%	54.8%
		% within dif_messagestyles	30.8%	24.3%	33.3%	0.0%	28.3%
	Formal	Count	9	28	2	7	43
		% within Formal_informal	10.1%	31.5%	2.2%	7.9%	48.3%
		% within dif_messagestyles	69.2%	75.7%	66.7%	100.0%	71.7%

The third research question concerned the relationship between the level of formality of each message style within the framework and the paralinguistic language features. To get an overview of the presence of paralinguistic language features in the tweets, a descriptive statistics frequencies test was executed. The results indicated that 110 (91.7%) of the tweets contained paralinguistic features, while 10 (8.3%) did not (*M* = 60.50, *SD* = 34.78).

Consequently, to identify the potential relationship between the variables, two Chi-square tests were executed. Firstly, the test was carried out for the variables 'paralinguistic language features' and 'message styles'. The results indicated, however, that there was no significant relationship between the paralinguistic language features and the message styles − no message style category included (χ^2 (4) = 4.45, *p* = .355). A Fischer's Exact test indicated a value of 3.93, *p* = .325. More results can be found in Table 5.

Table 5: Paralinguistic features * Message styles Crosstabulation

			Message styles				
			Symbolic	Emo.	Generic	Preemptive	No m.s
Paralinguistic features	Informal	Count	1	2	0	2	5
		% within Paralinguistic_features	10.0%	20.0 %	0.0%	20.0%	50.0%
		% within dif_messagestyles	7.7%	5.4%	0.0%	28.6%	8.3%
	Formal	Count	12	35	3	5	55
		% within Paralinguistic_features	10.9%	31.8 %	2.7%	4.5%	50.0%
		% within dif_messagestyles	92.3%	94.6 %	100.0%	71.4%	91.7%

Secondly, the relationship between the variables 'paralinguistic language features' and 'level of formality (formal/ informal)' was tested. Similarly, no significant relationship between the two variables was found (χ^2 (1) = .099, p = 1.000). A Fischer's Exact Test indicated a value of p = .717. In Table 6, further results are indicated.

Table 6: **Paralinguistic features * Level of formality Crosstabulation**

			Level of formality	
			Informal	Formal
Paralinguistic features	Informal	Count	3	7
		% within Paralinguistic_features	30.0%	70.0%
		% within Formal_informal	9.7%	7.9%
	Formal	Count	28	82
		% within Paralinguistic_features	25.5%	74.5%
		% within Formal_informal	90.3%	92.1%
	Total	Count	31	89
		% within Paralinguistic_features	25.8%	74.2%
		% within Formal_informal	100.0%	100.0%

Conclusion and discussion

The objective of this study was to determine whether the corporate communication message style framework developed by Cornelissen (2008) can be integrated on Twitter and whether companies can associate each component message style of the framework with the (in) formality level and with the paralinguistic language features, in order to improve their communication strategies.

To begin with, the analysis revealed that most of the tweets could not fit into any message style category. This might be due to the fact companies are not aware of the message style framework and communication through Twitter does not hold the same function as the message style framework does in the traditional, offline communication. As mentioned before, language on Twitter is fundamentally different than the language on the traditional media channels. Because of the character constraint, the messages (tweets) are much shorter – on average in this study approximately 6 words short – which makes it more difficult for companies/brands to incorporate a certain message style. However, the second most frequent message style was the emotional one, suggesting that tweets are language excerpts loaded with emotions, which is in line with Pak and Paroubek's (2010) conclusions. Furthermore, the results indicated Carlsberg posted tweets in 3 message styles present in the sample: emotional, generic, and preemptive. Considering Carlsberg is a beer brand, the classification of tweets into the emotional message style was rather unexpected, since evoking emotions might fit better with the consumption of chocolate, whether it is the Toblerone brand or a different one. However, consumption of beer is regarded as a very social activity and numerous beer brands, including Carlsberg, are endorsing or sponsoring sports events, especially soccer; because of its social character, consuming beer can evoke emotions such as excitement or joy. The majority of tweets that fit into the no the message style category was posted by LAY's. The explanation can be found in the fact LAY's organised a contest on their Twitter page, using simple and objective language structures that did not indicate affiliation with a message style. Consumption of chocolate has always been associated with a feeling of pleasure, which is why it was predictable that Toblerone would prefer a more metaphoric approach, resulting in a higher number of tweets fitting into the emotional and symbolic message styles.

Next, the calculation of the *F-score* revealed the majority of the tweets were formal. This is surprising, since language on Twitter was expected to be informal. Also, despite the non-existent relationship between the (in) formality level of each message style within the framework and the paralinguistic language features, companies from the FMCG industry use

the latter very often. The explanation can be found in the fact that those features are specific for the online communication, as stated by Schandorf (2013). Among the most frequent characteristics were the exclamation sign, letter capitalisation and the ellipsis ('…'). Surprisingly, very few tweets contained smileys, which is the opposite of Pak and Paroubek's (2010) findings, who revealed that emotions are frequently conveyed in tweets through happy and sad emoticons. Indeed, in this case, written messages were mostly formal, but the paralinguistic language features were prevalent. Paralinguistic language features are informal cues, which is why the results of this study were similar to the ones of Walther and Burgoon (1992). As it was seen in the results and throughout the analysis, a high level of formality did not necessarily mean that the language excerpt was written in a serious tone. For example, in the following tweet – 3*7 reasons why Denmark is awesome. Probably hard to argue with 29! No. 2 is Stockholm though, @BuzzFeed ;)* – Carlsberg uses a humorous, playful tone, completed by the emoticon ";)", but the calculation of the *F-score* revealed the tweet was formal. An important part of the tweets was found to be somewhat humoristic and involved suggestive, metaphoric language, fact which does not correspond with Irving's (1979) view, who suggests written communication is necessarily solemn.

Limitations and further research

The first flaws occurred during the coding process, when several constraints about the coding of the message styles and of the formality level occurred. More than often, the tweets were accompanied by (suggestive) photos and/or videos. As the current study focused only the language of the tweets, the photos and/or videos were omitted from the investigation. This choice might have biased the results in a slight manner, because if the photos and videos were not omitted, the association of a tweet with a message style could have been different. Also, since the tweets were so short, it would have been more expected that a message style would be contained in the photos or videos, rather than in the tweet as a text message.

Despite the similar nature of the brands, due to differences between the products they represent, dissimilarities between their Twitter communication strategies were also encountered. In the present study, they were recorded as limitations. Indeed, the selected brands are representative for the FMCG industry, but the products require different communication approaches. For example, in the case of LAY's, the company organised a contest for followers, so as a consequence, firstly an important part of the tweets did not fit in any message style and secondly, most of the tweets were repeated, thus limiting the chance of

using more distinct paralinguistic language. Perhaps if the part of sample from LAY's was from a different period, more tweets would have been of other message styles and would have contained more or different paralinguistic language features. The tweet sample and the number of brands investigated might not have been representative and sufficient, which, in turn, might have led to a limited applicability of the present study; the results may not be generalised to the whole FMCG industry.

When calculating the *F-score*, it was often the case that some component words of the tweets could not fit in any conventional part of speech category, so they were left out. Besides, in most of the cases, hashtags were taken as parts of speech based on their function in the tweets and also to be able to reach to realistic *F-scores*. A fairly notable limitation was the rather narrow and inaccurate programming of the online parts of speech detection tool; for instance, the software is not pre-set to detect interjections. Overall, it seems that the *F-score* lacks the accuracy to calculate a solid, realistic level of formality of language excerpts or perhaps it is rather outdated. The most striking examples of *F-score* will be provided in the following subsection. The explanation might be that when the *F-score* came out, obviously neither the internet, nor the social networks were existent.

The message style framework might not be known by companies/brands, despite being a part of the curriculum of the communication sciences studies or it is not a widespread communication tool. Companies might use it as part of their communication strategy, but without being aware of its formal name.

Since Twitter is the most popular social network and social media has taken over most of the face-to-face communication, scholars should focus more on researching the paralinguistic language in messages. At the moment there is hardly any solid literature on what paralinguistic language consists of on Twitter, as the microblog already has some particular features such as the hashtag (#), which might themselves be considered as paralinguistic language.

Focus on F-score limitations

The results of the study indicated that a number of tweets stood out as containing special features or being an exception from the coding rules, and they were treated as limitations. In the following paragraphs, an overview of the most exceptional tweets will be provided.

Tweet (65), LAY'S: Chipotle Mayo or Maple Syrup? Tell us what new flavor you want, you could win $1 million! See Rules http://bit.ly/DoUsAFlavor7 #DoUsAFlavor

Although followers' spirit of competition is being challenged, the tweet was classified under no message style because the language used was extremely objective. The calculation of the *F-score* revealed an interesting score of 50.005%, which decided the formality of the tweet by 0.005%. Yet, this might be a solid example why the *F-score* is not accurate enough to determine a realistic (in) formality level. The rules state that everything that is below 50% is informal and everything that is above 50% is formal, but can a difference of 0.005% actually make a difference? Also, the same coding rules do not indicate a special (in) formality level for the *F-scores* of 50%, which adds to the constraint of this case.

Tweet (12), Carlsberg: MARKOVIC!!!! #LIVTOT

The lack of information derived made it hard for this tweet to be placed under a category of message style. Only followers who are soccer fans will understand the tweet makes reference to a soccer match and to a particular soccer player. This tweet was the highest formal language excerpt of the whole sample, with a score of *F* of 100%. The pure formality was given by the 2 containing words, which were both nouns (the hashtag was taken as noun). However, this situation was treated as a limitation because Heylighen and Dewaele (2002) explain that an *F-score* will never reach the outer limit of 100%. Additionally, the researcher had to count the hashtag as a part of speech in order to be able to calculate the *F-score*.

Tweet (113), Toblerone: Beam me up choccy! #SciFiDayIn

The tweet was regarded as symbolic, due to the association of the chocolate brand with the 'Science-Fiction Day', implying that Toblerone actively supports global events and suggesting followers to do so as well. As mentioned in the Method section, out of coding consideration, the *F-scores* lower than 50% were regarded as informal and the ones higher than 50% as formal. However, during the calculation of the *F-score* it was sometimes the case that the *F-scores* obtained were 50% sharp. These cases were recorded as limitations and repaired by the researcher. Because an *F-score* of 50% contradicted the coding rules, the decision to place the tweets into either of the categories (formal or informal) was based on the proportions of the component parts of speech. As a result, the aforementioned tweet was categorised as informal since the proportions of pronouns and verbs were higher than those of nouns, adjectives, prepositions or articles. Out of 5 words contained by the tweet, 1 was a noun and 2 were a pronoun and a verb.

Finally, the present study should be seen as an encouragement for more research about

language on Twitter and the possible association with a more traditional corporate communication approach such as the ‘message style framework’. The current research is a solid proof of the originality of Twitter as a modern channel and an important contribution to the improvement of the use of social media in corporate communication. Organisations (brands) must find a way to merge the traditional channels with the contemporary approaches, such as social media – which is not only a trend anymore, but a vital necessity in corporate communication.

References

Bombelli, G., Soler, L.R., & Waasaf, M.S. (2013). The language of Evaluation: paralinguistic features as a phonological domain for appraisal. *DELTA*, *29*(2), 267-280.

Bortolini, U., C. Tagliavini & A. Zampolli (1971). *Lessico di frequenza dellalingua italiana contemporanea*, IBM Italia.

Brown, G. (1990). *Listening to spoken discourse*, 2nd ed., London: Longman.

Clow, K.E., & Baack, D. (2007). Integrated Advertising, Promotion, and Marketing Communications, 3rd ed., Upper Saddle River, N.J.: Pearson Prentice Hall.

Cornelissen, J. (2008). *Corporate Communication: A guide to theory and practice.* London: Sage Publications.

Dibb, S., Simkin, L., Pride, W., & Ferrell, O. (2006), *Marketing: Concepts and Strategies*, 5th ed., Boston, MA: Houghton Mifflin.

Dovey, F. (2009, August 10). FMCG brands must follow their customers online however hard it is. *New Media Age*, p15-15. 1/3p.

Ferrara, R, Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication*, *8*, 8-34.

Herring, S.C. (2004). Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. To appear in Barab, S. A., Kling, R., & Gray, J. H. (Eds.). Designing for Virtual Communities in the Service of Learning, 338-376, New York, NY: Cambridge University Press.

Gains, J. (1999). Electronic mail-a new style of communication or just a new medium? An investigation into the text features of e-mail. *English for Specific Purposes*, *18*, 81-101.

Gordon, I. (2002). Cultural Symbols. In McDonough, J., & Egolf, K. *The Advertising Age Encyclopedia of Advertising*, Chicago, IL: Fitzroy Dearborn.

Heylighen, F., & Dewaele, J.-M. (1999). Formality of Language: definition, measurement and behavioral determinants. Internal Report, Center "Leo Apostel", Free University of Brussels.

Heylighen, F. & Dewaele, J.-M. (2002). Variation in the contextuality of language: an empirical Measure. *Foundations of Science*, *7*(3), 293-340.

Hudson, R. (1994). About 37% of Word-tokens Are Nouns. *Language*, *70*, 331-339.

Irvine, J. (1979). Formality and Informality in Communicative Events. *American Anthropologist, 8* (4), 773-790.

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. Lund University, Dept. of Linguistics and Phonetics. *Working Papers, 53*, 61-79.

Juilland, A., & V. Traversa (1973). *Frequency Dictionary of Italian Words*. The Hague: Mouton.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lovejoy, K., Waters, R. D., & Saxton, G. D. (2012). Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. *Public Relations Review*, *38*(2), 313-318.

Messner, M., Jin, Y., Medina-Messner, V., Meganck, S., Quarforth, S.C., & Norton, S.K. (2013). 140 Characters for Better Health: An Exploration of the Twitter Engagement of Leading Nonprofit Organizations. In H.S. Noor Al-Deen & J.A. Hendricks (Ed.). *Social Media and Strategic Communications*, Palgrave Macmillan.

Moore, D.J., Harris, W.D., & Chen, H.C. (1995), Affect intensity: an individual difference response to advertising appeals. *Journal of Consumer Research, 22* (2), p. 154-164.

Mosquera, A., & Moreda, P. (2011). The Use of Metrics for Measuring Informality Levels in Web 2.0 Texts. *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology*, 184-188, Brazil.

Muralidharan, S., Rasmussen, L., Patterson, D., & Shin, J.-H. (2011). Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts. *Public Relations Review, 37*(2), 175-177.

Murray, D. E. (1991). Conversation for action: The computer terminal as medium of communication. Amsterdam: Benjamins.

Murray, D. E. (2000). Protean Communication: The Language of Computer-Mediated Communication. *TESOL Quarterly*, *34*(3), TESOL in the 21st Century, 397-421.

Norris, A.R. (2012). Computer-Mediated Communication and Globalization: Considering Social, Academic, and Business Factors. *Studentpulse.com*, *4*(2), 1, United States of America.

Pak, A. & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining, France.

Richards, J.C., Platt, J., & Platt, H. (1997). *Dictionary of Language Teaching and Applied Linguistics*, London: Longman.

Schandorf, M. (2013). Mediated gesture: Paralinguistic communication and phatic text. *Convergence: The International Journal of Research into New Media Technologies, 19*(3), 319-344.

Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., & Narayanan, S. (2013). Paralinguistics in speech and language − State-of-the-art and the challenge. *Computer Speech and Language*, *27*, 4-39.

Smith, B. (2003). The use of communication strategies in computer-mediated communication. *System*, *31*(1), 29-53. doi: 10.1016/S0346-251X(02)00072-6.

Stelzner, M.A. (2009). Social media marketing industry report. Retrieved on 12 April 2015 from http://www.socialmediasummit09.com/.

Tench, P. (1990). *The roles of intonation in English discourse*, Frankfurt am Main: Verlag Peter Lang GMBH.

Thurlow, C., Lengel, L., & Tomic, A. (2004). *Computer Mediated Communication: Social Interaction and the Internet*, London: Sage.

Walther, B. J., & Burgoon, K. J. (1992). Relational Communication in Computer-Mediated Interaction. International Communication Research, *Human Communication Research*, *19*(1), 50-88.

Waters, R.D., & Jamal, J.Y. (2011). Tweet, tweet, tweet: A content analysis of nonprofit organizations' Twitter updates. *Public Relations Review, 37,* 321-324.

Yates, J., & Orlikowski, W.J. (1993). Knee-jerk anti-LOOPism and other e-mail phenomena: Oral, written and electronic patterns in computer-mediated communication (*MIT Sloan School Working Paper*, No. 3578-93). Retrieved September 22, 2015, from http://ccs.mit.edu/papers/CCSWP150.html.

Zhang, H., Sun, J., Liu, F., & Knight, J.G. (2014). Be rational or be emotional: advertising appeals, service types and consumer responses. *European Journal of Marketing, 48*(11/12), 2105-2126.

Appendix 1

This appendix provides further examples of tweets belonging to all the message styles with different levels of formality and with a detailed description of the paralinguistic language features they contain.

Table 7: Illustrations of all the message styles in tweets, formality level and paralinguistic / phatic language features

Tweet no.	*Tweet*	*Company*	*Message style*	*Formality level*	*Paralinguistic / phatic language features*
1	The only right way to use the last page of your magazine. Probably. #DIY http://fal.cn/probably	Carlsberg	Emotional	Formal (F = 67.85%)	DYI – special capitalisation of letters that resulted in a worldwide known acronym for "Do It Yourself"
7	Probably the only right gift. Carlsberg ad, 1936. #TBT #ThrowbackThursday	Carlsberg	Preemptive	Formal (F = 75%)	TBT = slang word / abbreviation / acronym for Throwback Thursday, often used in social media to point out an old photo or idea
10	Crown caps where invented in 1892 & so was that great feeling of opening a cold beer. Definitely worth a Nobel. #TBT	Carlsberg	Generic	Formal (F = 68.18%)	TBT = slang word / abbreviation / acronym for Throwback Thursday, often used in social media to point

					out an old photo or idea
30	CONGRATS @VinnieJones! You won 'Best Actor' #IfCarlsbergDidAwards #BestActor	Carlsberg	No message style	Formal (F = 62.5%)	Special capitalisation: B, A, I, C, D, A, B, C to emphasise the importance and enthusiasm of the event
41	Treat Yourself Tuesday: Turn a caprese salad into a dip & scoop it up with your fave chips. Voila, the perfect snack!	LAY'S	Emotional	Formal (F = 65.90%)	Special capitalisation: T, Y, T that draws attention upon the prominence of the occasion. Also, the interjection 'Voila' together with the exclamation mark are vocal gestures
43	Lay's Wavy is the new pita. Take your taste buds overseas with this yummy Mediterranean plate	LAY'S	Preemptive	Formal (F = 74.87%)	No paralinguistic / phatic language features present
61	Get inspired by the Pacific Northwest w/@AshleyRodriguez this week on #LaysInstaTrip! http://Instagram.com/AshRod	LAY'S	Symbolic	Formal (F = 79.16%)	Special capitalisation: L, I, T; !
85	We're going big with the ultimate secret weapon… #AntMan	Toblerone	Symbolic	Formal (F = 60%)	Ellipsis: …
110	Tell one, tell all! We have a VERY special #ToblerоneTuesday coming at 12... see you then!	Toblerone	Emotional	Informal (F = 26.04%)	Special capitalisation: VERY, T, ellipsis (…), s(ee), !
117	Think we know who number 1 is… RT @BuzzFeed: Can You Identify	Toblerone	Generic	Informal (F =	Ellipsis: …

	10 Iconic Chocolates? http://bzfd.it/1yZ6Wf0			40.62%)	
118	Nice easy one this week, simply RT & follow to win a Toblerone of your choice! #ToblerоneTuesday	Toblerone	No message style	Formal (F = 64.70%)	!, special capitalisation