A COMPUTATIONAL ANALYSIS OF THE MORPHOSYNTACTIC VARIATION IN TWEETS WRITTEN BY SPANISH JOURNALISTS

ESTEBAN VÁZQUEZ-CANO[1], M.ª LUISA SEVILLANO[2], JOSÉ MANUEL SÁEZ-LÓPEZ[3]

Abstract. This article presents a research that analysed variations in the linguistic construction of tweets from a sample of 10 of Spain's most active and widely followed journalists on the microblogging site Twitter. We applied a methodology based on text mining and framed by tokenization, lemmatization and morphosyntactic tagging in order to analyse the main characteristics and variations in the journalists' tweets in terms of word class and the most recurrent syntactic functions and linguistic structures. The results show that nouns, prepositions and verbs are the words most widely used, with the dominant functions being the direct object and the attribute, both of which characterise conceptual and argumentative discourse by two types of linguistic patterns: the noun and adverbial subordinate clauses.

Key words: digital journalism, tweet, Twitter, morphosyntax, computational linguistics, digital language.

1. INTRODUCTION

The incessant use of social networks, instant messaging apps and microblogging is generating a new form of digital writing that permeates all aspects of life, from the personal and social to the academic and professional. The microblogging site Twitter has emerged as a significant force among the public at large, and in journalism.in particular (Ahmad, 2010; Broersma & Graham, 2013; Lasorsa, Lewis, & Holton, 2012). The bidirectional flow of information and the reaction that content causes on the social networks encourages constant interaction, and digital writing has had to adapt to the new audiovisual and digital scenario, and condense ideas into 140 characters on networks such as Twitter (Honeycutt & Herring, 2009; Hong, Convertino, & Chi, 2011; Pano-Alamán, & Mancera-Rueda, 2014). These flows are disrupted by the presence of the audiovisual narrative, with the application of forceful communication strategies with their origin in the media and the social networks.
The language of journalism has been affected by this social network activity that has generated a new way of presenting the journalistic message through a new digital discourse

[1] Faculty of Education. Universidad Nacional de Educación a Distancia. Madrid (UNED), evazquez@edu.uned.es.
[2] Faculty of Education. Universidad Nacional de Educación a Distancia. Madrid (UNED), mlsevillano@edu.uned.es.
[3] Faculty of Education. Universidad Nacional de Educación a Distancia, Madrid (UNED), jmsaezlopez@edu.uned.es.

RRL, **LXV**, *4*, p. 331–349, Bucureşti, 2020

that significantly alters syntactic construction, lexical selection and orthotypography; and it incorporates paralinguistic elements (iconic-visual) that enrich and modulate opinion, the news and journalistic argumentation. This study examines the digital contributions of journalists on Twitter using a methodology based on text mining and computational and statistical analysis, with parameters such as tokenization and lemmatization, and the significance and incidence of formal variables in the construction and variation of journalistic tweets. Furthermore, we compare the morphosyntax of written Spanish in Twitter with the absolute frequency of word classes in standard written Spanish according to the Spanish Corpus of the XXI Century (RAE, 2015) and the syntactic patterns with respect to the average frequency documented in Spanish written language (Syntactic Database of the current Spanish, 2001).

2. TWITTER LANGUAGE

Since Jack Dorsey set up Twitter in 2006, it has become a worldwide phenomenon. Its users can communicate in real time, follow other users and see what they are up to, and interact with them via 140-character messages (Java et al., 2007). Twitter, as a social and technological phenomenon, is half-way between a social network and an instant messaging service, which has led to the creation of a code of communication and specific guidelines for interaction (Mancera-Rueda & Pano-Alamán, 2013). A survey, "Join the Conversation: How Spanish Journalists are using Twitter" (Carrera-Álvarez et al., 2012), carried out by Journalism students at the Universidad Carlos III in Madrid (Spain) concluded that Twitter is one of the social networks most widely respected by journalists.

Today digital writing is a multifaceted concept; it appears in all manner of situations and on numerous devices, and in multiple digital contexts that are personal, academic or professional, in which citizens express themselves. Technological interaction has generated a form of writing that is ubiquitous (Bodomo, 2009; Vázquez-Cano, 2012; 2015; Vázquez-Cano, López-Meneses, & Sevillano, 2017), and in journalism this has produced a kind of "early network alert" system with the rebirth of flash journalism (Carrión, 2013). This new context has obliged journalists to adapt their styles and editorial techniques to the requirements of these new channels of communication, where orthotypographic and paralinguistic elements acquire a new dimension and news-providing function (Thurlow & Poff, 2011). The internal elements of the tweet, such as hashtags and links to other news stories on the Net, help enrich the story and can take on new interpretative forms. So, the use of orthotypographic alterations, emoticons or the insertion of audiovisual elements (photos, audio, images, illustrations, montages, memes, etc.) can greatly enhance the journalist's communicative intent from within parameters that are completely different from those used in more traditional journalism media.

Twitter is synonymous with communicative immediacy and a mix of private and public communication. The linguistic implications of this phenomenon affect different aspects of language, namely orthotypography, morphology, syntax, semantics and pragmatics. In this digital, synchronic and ubiquitous exchange it is the interaction, not the narrative, that is controlled, and this produces a "conversational ecology" (Boyd, Golder, & Lotan, 2010; Markman, 2013). Authorship is shared, as the participants' successive interventions facilitate the continuity of the story which is, by its very nature, improvised and spontaneous, with under-developed syntax and, sometimes, "relaxed" orthography

(Gómez-Camacho, 2007; 2014). For example, in Spanish: confusion in phonemes and the corresponding letters: "haya"(verb "to be", subjuctive in Spanish)and "allá" (adverb of place) or shortenings in basics words as "q" instead of "que" (relative pronoum) or others alteration which do not correspond to any communicative process, as writing "Kasa" instead of "casa" (/k/ is the phoneme to represent three letters in Spanish "c" "k" and "qu". This relaxing and confusion of letters and phonemesis due to a relaxed attitude in some cases, but in others one is the evidence lack of spelling knowledge. Research into the use of language on Twitter and into subgenres like Twitter Journalism is made easier on this microblogging site than on other social networks because Twitter contains forms and content that are more stable (Cortés-Rodríguez, 2012; Lomborg, 2011).

3. UBIQUITOUS DIGITAL WRITING AND "TWITTER-LANGUAGE"

The main linguistic features of digital communication are vowel and consonant lengthening, the use of emoticons, a laid-back attitude towards spelling and hypertextuality. In principle, these should hinder the correct pragmatic reading of the text, yet communication in the digital setting does not break down because the interlocutors adopt efficient strategies to reconstruct the social conversation pathways and contextual information in ways typical of face-to-face interaction by, for example, reproducing suprasegmental elements within the written text (Mancera-Rueda & Pano-Alamán, 2013: 12). In Twitter, communicative immediacy prevails, and private and public communication are interlinked; there is greater emotional involvement, and messages are contextualised in a specific communicative situation that provides the reader with an interpretative intertextual story. After reading more than 10 tweets on the same trending topic, the user of the network can construct a discursive micro-story of the situation initiated or under discussion. This 140-character discourse approximates to a chat-type dialogue situation. In this digital, synchronic and ubiquitous exchange, it is not the narrative but the interaction that is controlled (Markman, 2013). Authorship is shared, as it is the interlocutors' successive interventions that build continuity in the story, which is usually improvised, rarely thought through, with basic syntax and a haphazard application of spelling rules in the informal context (Gómez-Camacho, 2014).

One of the most important aspects of the new network languages is not so much the technology as the fact that words have become public property. One of the main changes is public writing. Looking beyond the merely technological,it is interesting to see how the written word has now gone public. This is the real novelty for people and companies. When somebody writes something for the public, and it is read by more people than the author realises, writers start holding themselves and the writing of others to much higher standards. Given that writing in the digital medium cannot incorporate modulatory elements typical of oral discourse, it compensates with the use of orthotypographic and paralinguistic elements such as emoticons and, especially, capital letters to express surprise or to emphasise a particular piece of information. To enrich tweet content, many tweets come attached with links to websites, videos and images recommended by followers. However, since such links can be very long and take up too many characters, URL shorteners are used to generate a unique abbreviated web address. Another way of enriching tweet content is to include a label or hashtag, with the # sign followed by a word or syntagma to indicate the subject of the tweet. The labels represent explicit metadata on the content mediated by a tweet, and as such form part of the linguistic structure (Menna, 2012).

The concept of journalistic language as a specialized register has been widely described in journalism (Hernando, 1990:44); it is seen as heterogeneous, and characterised by a wide range of journalistic subgenres and news reporting media. Lázaro Carreter (1977:10) examined this permeability of journalistic language in detail, with its array of discursive variables, and established that besides the prevailing standard style, it also contained elements of other registers that include literature, the legal and administrative and the colloquial. "These three frontiers mark out a space in which, I believe, newspaper language should operate with complete ease". For example, social networks force journalists to improve constantly; Twitter obliges them to demonstrate their worthiness to be called a journalist (Mancera-Rueda & Pano-Alamán, 2013).

On Twitter, the imposition of characters limitation is key to understanding the transformation of language. Abbreviations have always existed and always acted as a shortcut for people sharing a common code; but they often seemed like errors to those outside that code. Expressions such as "jejeje, jajaja, tqm, tons, LOL", etc. became supralanguages that enabled interaction between young tweeters for whom these made absolute sense through their sheer frequency of use; but not for others, who often fail to understand the messages they receive. This type of digital communication, made easier by the cybermedia context described, provides the written register with features that are typical of an oral register. It does this by trying to make writing resemble as far as possible a spoken conversation, sometimes within the "group of friends" format, as the scientific literature testifies: "oralized written text", (Crystal, 2008), "oralized writing" (Dresner & Herring, 2012; Fortunati, 2001; Gómez-Camacho, 2007), "written orality" (Jaffe & Walton, 2000; Hutchby & Tanna, 2008) or "written conversations" (Fraca de Barrera, 2006).

Ubiquitous digital conversations demand interaction and reciprocity just as in face-to-face conversation, with simultaneity the dominant feature of both (Horstmansh of & Power, 2005; Lewis & Fabos, 2005; Riordan, Markman, & Stewart, 2013). The linguistic implications affect the orthotypographic, morphological, syntactic, semantic and pragmatic areas of language. The analysis of digital writing has profound repercussions for sociolinguistics and the parameters of the Information and Communication Society. It can help explain the synchrony of language as it constantly adapts to digital media, and can delineate its most important characteristics in order for us to understand its usage in different contexts such as that of journalism (Yus, 2001; Thurlow & Mroczek, 2011; Vázquez-Cano, 2012; Vázquez-Cano, Fombona, & Bernal, 2016).

So far, there have been no morphosyntactic studies of the Spanish language used by any subgenre on Twitter, so we have no models with which to compare our results. There are several studies on the English and Japanese used in this context (Bessho, Harada, & Kuniyoshi, 2012; Yoshino, Mori, & Kawahara, 2011; Inaba, Kamizono, & Takahashi, 2013; Sugiyama et al., 2013), but they do not allow us to engage in a formal comparison since both languages have a morphosyntactic structure that differs greatly from Spanish.

4. METHOD

The aim of this article is to establish the morphosyntactic features of the linguistic construction of journalistic tweets and so determine patterns of use of the language used by a sample of Spanish journalists on Twitter. The parameters were tokenization, lemmatization and grammatical tagging, to which we applied a computational and statistical treatment. To achieve this, we adopted a research methodology that formed part

of a computer-mediated discourse analysis using Computational Linguistics for Text Analysis techniques (Fletcher, 2004; Parodi, 2010; Vázquez-Cano, Mengual, & Roig, 2015, Vázquez-Cano, Fombona, & Bernal, 2016), as well as statistical inference processing in the analysis of the linguistic construction of the digital message. The linguistic analysis consisted of four phases: I) extraction of the journalists' tweets and metrics on Twitter; II) identification of tokenization, lemmatization and grammatical tagging; III) a descriptive, inferential and statistical analysis of the asymmetry of the linguistic construction, and parametric tests by means of the relation of simple multiple regression analyses. These parametric tests allow us to analyse lexical densities of part of the speech (noun, verb, adjective, …) and to determine the more significant syntactic patterns. This was done in order to check for the possible influence of the study variables on the linguistic characterization of the journalists' tweets. We also ran non-parametric Mann-Whitney U tests to determine the influence of gender on the sample.

In the first phase (I), we used the "Tweet Chup" tool (http://tweetchup.com) to analyse profile metrics on Twitter during a specific period (during two weeks). After, we exported the tweets to Excel to generate an .xls file that we could use for analysis with text mining and inferential statistics techniques in the second phase of the methodological procedure.

In the development of the second phase (II) we used Meaning Cloud's API (Lemmatization, PoS and Parsing Console) text mining tool. We used automated algorithms for tokenization (tweet segmentation) and lemmatization (seme disambiguation) to identify the most important morphosyntactic elements. The procedure to obtain the "tokenized" text corresponds to a mathematical structure like this: t_i = ("$w_{1\oplus\boxtimes}w_{2\oplus\boxtimes}\ldots..\oplus w_{m+\boxtimes}$"), where t_i is an "i^{th} tweet" chain with a fixed number "m" of words, operator "$\oplus$" marks the words separated by spaces, and operator "$\boxtimes$" marks the paralinguistic element separated by pausesin order to calculate its frequency in the written texts. Then a "PoS Tagging" (part-of-speech tagging / grammatical tagging) is applied. Using this technique enabled us to assign or tag each word in the tweets, according to its grammatical category, based on the definition of the word and on the context in which it appeared (that is, the relation to adjacent and related words in a sentence or paragraph). Finally, we applied the API Meaning Cloud linguistic algorithm based on the "Spanish Resource Grammar: HPSG (Head-Driven Phrase Structure Grammar) used by LKB (Linguistic Knowledge Building) for rules-based grammatical tagging; this enabled us to determine the underlying syntactic structure in each tweet based on the parameters of word class, syntactic function and linguistic and phrasal structure.

In the third phase, we used the SPSS 19 program to run a descriptive statistical analysis of the results obtained, and we analysed the tweets' "asymmetry" and "kurtosis" to outline their potentially recurring structures. The representation of the syntactic structure obtained was analysed by applying the grammatical tagging technique based on syntactic trees. We also examined whether gender produced significant differences.

Sample and selection

We based the selection of the 10 journalists from Spain on the following criteria: number of followers, average number of tweets posted per day and the number of times these were retweeted. We selected 10 Spanish journalists, five men and five women. Table 1 shows the activity data and tweet repercussion for each journalist during a random period of 15 days, between 28 May and 11 June 2015.

Table 1

The journalists and their activity on Twitter during two weeks
(May 26th–June 9th. See Figure 1)

Journalist	*Nº Followers*	*Nº Tweets in the 15-day period*	*Average Nº tweets per day*	*Nº retweets*
Pedro J. Ramírez	342465	464	33,1	8451
Juan Ramón Lucas	215958	107	7,6	2488
Ignacio Escolar	500134	297	21,2	51810
Jesús Maraña	138059	191	13,6	12376
Melchor Miralles	103578	135	9,6	371
Pepa Bueno	140195	268	19,1	24301
Ana Pastor	1334927	1068	76,3	104015
Susana Griso	403669	147	10,5	6200
Esther Palomera	57120	116	8,3	3350
Julia Otero	529178	108	7,7	8227
Media	37652,83	290,1	20,7	22158,9
Total	3765283	2901	207	221589

The tweets analysed are shown in Figure 1, which indicates the journalist who sent the tweet, and the date and time it was posted.

Pedro J. Ramírez	22:48 - 27 de may./22:52 - 28 de may./14:04 - 7 de jun. de 2015/22:36 - 10 de jun./13:05 - 29 de may./22:38 - 1 de jun./11:26 - 1 de jun./22:40 - 2 de jun./22:47 - 9 de jun./0:06 - 2 de jun./22:52 - 3 de jun./ 22:48 - 4 de jun. /0:21 - 7 de jun./6:25 - 6 de jun./13:47 - 29 de may./1:52 - 29 de may./1:01 - 31 de may./13:06 - 28 de may./23:01 - 28 de may.
Juan R. Lucas	3:12 - 30 de may./10:43 - 30 de may./14:24 - 6 de jun./9:24 - 30 de may./22:11 - 9 de jun./14:16 - 6 de jun./0:09 - 4 de jun./4:16 - 31 de may./3:46 - 6 de jun./0:38 - 31 de may./10:27 - 29 de may./2:04 - 7 de jun./7:35 - 7 de jun./10:19 - 30 de may./22:29 - 1 de jun.
Ignacio Escolar	9:49 - 28 de may./4:12 - 29 de may./8:36 - 5 de jun./8:25 - 28 de may./9:57 - 30 de may./4:26 - 28 de may./4:52 - 9 de jun./2:57 - 2 de jun./0:48 - 6 de jun./13:52 - 9 de jun./0:02 - 9 de jun./2:49 - 29 de may./2:23 - 28 de may./3:25 - 29 de may./3:50 - 3 de jun./11:52 - 1 de jun./10:02 - 30 de may./5:57 - 5 de jun.
Jesús Maraña	14:15 - 6 de jun./7:58 - 11 de jun./6:27 - 29 de may./10:56 - 2 de jun./2:56 - 1 de jun./14:54 - 5 de jun./1:13 - 3 de jun./9:20 - 8 de jun./15:03 - 3 de jun./1:52 - 4 de jun./2:51 - 2 de jun./1:19 - 3 de jun./1:38 - 2 de jun./14:27 - 1 de jun.
Melchor Miralles	15:28 - 3 de jun./3:48 - 28 de may./23:34 - 7 de jun./9:41 - 31 de may./6:13 - 9 de jun./14:42 - 30 de may./2:09 - 31 de may./15:49 - 28 de may./18:35 - 4 de jun./13:17 - 29 de may./2:08 - 31 de may./23:31 - 8 de jun./9:14 - 4 de jun./12:14 - 7 de jun./15:27 - 3 de jun./16:38 - 9 de jun.
Pepa Bueno	14:21 - 7 de jun./13:48 - 31 de may./14:04 - 31 de may./13:46 - 31 de may./14:08 - 31 de may./12:42 - 31 de may./14:33 - 7 de jun./13:29 - 31 de may./13:12 - 31 de may./13:28 - 7 de jun./11:30 - 7 de jun./13:01 - 31 de may./13:38 - 31 de may./13:33 - 30 de may./14:16 - 7 de jun./13:01 - 7 de jun./14:03 - 31 de may./14:33 - 7 de jun./13:36 - 31 de may.
Ana Pastor	5:01 - 30 de may./6:53 - 1 de jun./11:41 - 3 de jun./13:13 - 9 de jun./7:42 - 2 de jun./12:29 - 9 de jun./3:32 - 9 de jun./6:24 - 29 de may./13:37 - 1 de jun./13:32 - 7 de jun./13:29 - 8 de jun./0:48 - 10 de jun./4:15 - 4 de jun. 0:57 - 9 de jun./4:19 - 1 de jun./1:56 - 28 de may.
Susana Griso	5:11 - 9 de jun./4:20 - 4 de jun./5:28 - 29 de may./3:53 - 29 de may./5:04 - 9 de jun./4:47 - 11 de jun./1:48 - 4 de jun./1:51 - 11 de jun./1:56 - 4 de jun./1:41 - 29 de may./3:47 - 9 de jun./3:32 - 4 de jun./1:47 - 8 de jun./3:15 - 4 de jun./3:47 - 9 de jun./1:43 - 10 de jun./4:26 - 1 de jun.
Esther Palomera	6:07 - 11 de jun./23:38 - 1 de jun./22:13 - 7 de jun./13:29 - 9 de jun./22:23 - 2 de jun./8:28 - 9 de jun./3:45 - 2 de jun./2:07 - 30 de may./7:22 - 3 de jun./2:09 - 30 de may./0:10 - 31 de may./12:38 - 28 de may./22:48 - 7 de jun./23:44 - 3 de jun./22:34 - 31 de may./1:12 - 29 de may./11:16 - 2 de jun.
Julia Otero	15:50 - 8 de jun./12:41 - 7 de jun./15:55 - 8 de jun./12:37 - 7 de jun./5:16 - 5 de jun./16:00 - 8 de jun./3:16 - 3 de jun./14:24 - 7 de jun./14:40 - 31 de may./0:02 - 9 de jun./11:33 - 7 de jun./14:09 - 28 de may./14:06 - 31 de may./11:29 - 7 de jun./14:16 - 7 de jun./5:09 - 5 de jun./13:12 - 7 de jun.

Fig. 1. *Data analysed.*

Results and Discussion

The first results presented here relate to the construction of the tweet in terms of main linguistic characteristics (total number of words and characters, number of different words,

the average of words per clause, the number of sentences written and the word class used in them). We analysed each tweet with a view to tokenization, lemmatization and grammatical tagging in the output of the 10 journalists sampled (according to the word classes most widely used. Table 2 presents the descriptive statistics for the density of word classes in the tweets.

Table 2

Formal and linguistic characterization of the sample

Journalist and tweets	*Total number of words*	*Total number of characters*	*N° of differentwords*	*Average % of words per clause*	*N° sentenceswritten*
Pedro J. Ramírez (18)	419	2018	238	11,82	40
Juan R. Lucas	181	790	120	7,6	30
Ignacio Escolar	291	1874	190	10,13	33
Jesús Maraña	252	1284	126	10,26	21
Melchor Miralles	180	825	105	8,87	8
Pepa Bueno	280	1337	180	8,36	18
Ana Pastor	265	1272	159	7,66	38
Susana Griso	210	1045	137	7,83	30
Esther Palomera	192	993	131	10.14	21
Julia Otero	253	1272	156	7,89	35
Media	252,3	1271	154,2	8,042	27,4
Total	2523	12710	1542	80,42	274

As Table 2 shows, both male and female journalists use a similar total number of words (male = 1,323; female = 1,200). Likewise, the number of clauses in the set of tweets is very similar (male = 132; female = 142), as is the use of different words (male = 779; female = 763). This demonstrates that the difference between male and female journalists in terms of the formal use of the tweet in these categories is very similar and there are no significant differences. In contrast, differences between individual journalists are significant. The journalist with the highest number of words and sentences written is "Pedro J. Ramírez" with 419 words and 40 sentences, with an average word density per clause of 11.82%. The two journalists who use fewest words and sentences are Melchor Miralles (180 words in 8 sentences, with an average word density per clause of 8.87%) and Juan Ramón Lucas (181 words in 30 sentences, with an average word density per clause of 7.6%).These results show that the average number of words in the journalistic tweets analysed is similar to those found in generalist tweets (Hu, Talamadupula and Kambhampati, 2013), but a higher frequency than in the old SMS (Ling and Baron, 2007).
Firstly, the tokenization process enabled us to exclude words that were incomprehensible, then to eliminate spaces (represented by $\oplus$) to obtain the chain of orthotypographic and paralinguistic elements (represented by $\boxtimes$) that constitute tweets according to the different formulas, such as:t_i = ("$w_{1\oplus\boxtimes}w_{2\oplus\boxtimes}\ldots..\oplus w_{m+\boxtimes}$") (Table 3).

Table 3

Results of the tokenization processing of word classes.

	Noun	*Verb*	*Adjetive*	*Adverb*	*Preposition*	*Artícle*	*Pronoun*	*Conjunction*
N Valid	10	10	10	10	10	10	10	10
Lost	0	0	0	0	0	0	0	0
Mean	71.60	33.60	13.70	10.60	41.00	31.00	9.40	16.20
Median	66.50	29.50	11.50	11.00	37.50	31.00	6.50	12.00
St. Deviation	**25.334**	**11.890**	**7.288**	**3.169**	**12.632**	**8.832**	**7.260**	**9.555**
Variance	641.822	141.378	53.122	10.044	159.556	78.000	52.711	91.289
Asymmetry	.830	1.084	.898	.040	1.341	.242	.766	1.854
Kurtosis	**.314**	**.231**	**.244**	**–.507**	**1.520**	**.668**	**–.525**	**3.171**
Range	80	36	23	10	40	31	22	30
Minimum	43	21	5	6	29	17	1	9
Maximum	123	57	28	16	69	48	23	39
Total	716	336	137	106	410	310	94	162

Firstly, in Table 3, we present the descriptive results of the sample and the analysis of the standard deviation which tells us how spread out the presence of words and if they are concentrated around the mean or scattered. A total of 2271 words were tagged. Of the 2523 words posted, 252 were tags (hashtags) and “at” signs not codified in the linguistic study. The results show that the noun is the word most widely used in the tweets (n = 716/ 31.52%), followed by the preposition (n = 410 / 18.05%) and the verb (n = 336 / 14.79%). The absolute frequency of word classes in Spanish written according to the Spanish Corpus of the XXI Century (RAE, 2015) confirms that the noun is the most frequently word class used in written Spanish in Spain (Normalized frequency: 77,240,76 cases per million). The variation occurs in the use of the preposition and the verb. In tweets, the use of the preposition is more frequent than the verb. In standard written Spanish (Spain) the verb is more frequent (Normalized frequency: 49,693.03 cases per million) than the preposition (Normalized frequency: 38,646.97 cases per million).

The pre-eminence of the nounindicates that many tweets focus intransmitting information and this can help to keep a topical focus over time. This prevalence also favours to express the opinion (abstract, concrete or joke), as mentioned in previous studies (Wang, Chen, &Kan, 2016). The analysis of the asymmetry and kurtosis of a linguistic corpus enables us to identify whether the data are distributed uniformly around the mean. In the sample, the results show that the mean is greater than the median, and this generates an asymmetry and positive kurtosis (leptokurtic) in the categories of “noun”, “verb”, “adjective”, “preposition”, “article” and “conjunction”. Therefore, there is a high degree of concentration around the variable’s central values; consequently, these word classes are the most widely used by the 10 journalists in their tweets. In contrast, the kurtosis is negative for “pronoun” and “adverb”, which generates a platykurtic distribution, in other words, with less concentration around the central values of the distribution, meaning that these word classes are not so widely used in the tweets.

After, we ran successive simple linear regression tests to define the significance of each word class according to the total number of characters per tweet. We were able to determine the influence of each word class with respect to the formal structure of a tweet, which, of course, does not allow for more than 140 characters.

Table 4

Linear regression testing (word classes).

Model	*R*	*R-squared*	*Adjusted R-squared*	*Standard error of the estimate*	*Changes in thestatistics*				
					Change in R-squared	*Change in F*	*gl1*	*gl2*	*Significantchange in F*
Noun	.893	.797	.772	12.109	.797	31.394	1	8	.001
Verb	.859	.738	.705	6.454	.738	22.542	1	8	.001
Preposition	.919	.844	.824	5.295	.844	43.224	1	8	.000
Article	.844	.712	.676	5.028	.712	19.764	1	8	.002
Conjunction	.845	.714	.678	5.424	.714	19.930	1	8	.002

As we can observe, there are five significant word classes each with the following R^2: preposition (.824); noun (.772); verb (.705); conjunction (.678) and article (.676). This indicates that these word classes are the most frequently used when building a tweet. Later, we carried out successive multiple regression analyses to determine the potential word classes that are vital for building a common tweet (Table 5).

Table 5

Model for words classes.

Model		*Unstandarized Coefficients*		*Standarized Coefficients*	*t*	*Sig.*	*Correlations*			*Collinearity Statistics*	
		B	*Std. Error*	*Beta*			*Zero-Order*	*Partial*	*Part*	*Tol.*	*VIF*
1	(Const.)	39.248	33.762		1.163	.279					
	Prep.	5.196	.790	.919	6.574	.000	.919	.919	.919	1.000	1.000
2	(Const.)	14.770	16.993		.869	.414					
	Prep.	3.533	.498	.625	7.095	.000	.919	.937	.480	.590	1.696
	Verb	2.758	.529	.459	5.214	.001	.859	.892	.352	.590	1.696

The results from the multiple linear regression analyses show that only two word classes, “preposition” and “verb”, recur in the construction of tweets. These two word classes are the most common, regardless of the length of the tweet. Figure 2 shows the construction of an average tweet written by the journalists in this sample, with regard to its configuration and the word classes most widely used.

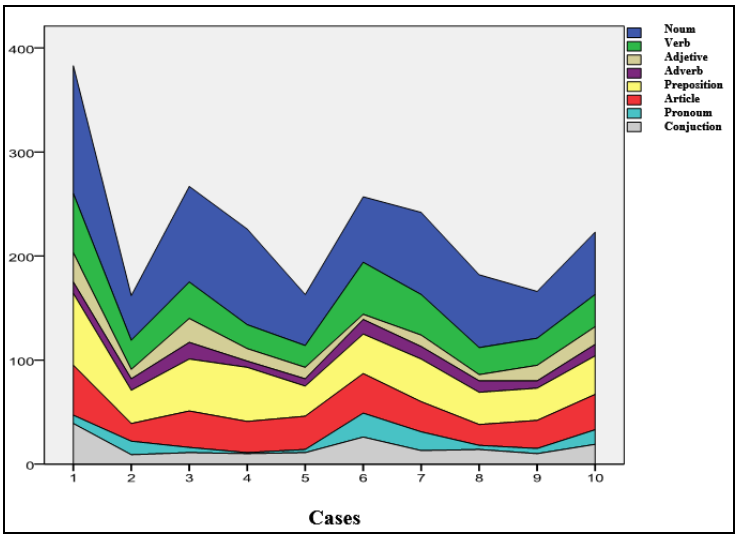

Fig. 2. *The words classes most widely used in the tweets.*

The "noun" and "preposition" are the word classes that register the highest percentage rate of appearances – mainly in the longer tweets – although the "verb" and "preposition" are the most homogenous word classes found in all tweet types, regardless of length. Another relevant aspect in the analysis of the construction of the journalistic tweet is to define the most important syntactic functions.

This analysis generates three types of linguistic structures in Spanish: attributive, transitive and intransitive. Demarcating the tweet's structure yields relevant information on the intention underlying the tweet's construction and the information transmitted ("attributive", based on aspects of argumentative discourse, "transitive", which focuses on the expression of ideas and concepts, and "intransitive", which refers to actions). Table 6 presents examples of tweets with transitive structures with direct objects and copulative structures with attributes. In this table, we can see the prevalence of the direct object, in green, over the other functions. While the distribution of the other functions is homogenous and tends to concentrate around a specific number of the tweets analysed, the direct object is distributed heterogeneously throughout the sample (Figure 3).

Table 6

Tweets with transitive and attributive patterns.

Journalists	*Tweets*	*T/A*[1]
PJR	Bdías. Si C's ha apoyado ***al PSOE*** en Andalucía y C's va a apoyar ***al PP*** en Madrid ¿por qué PSOE y PP no apoyan ***a C's*** en la ciudad de Valencia? *Gmorning. If C's has supported the PSOE in Andalusia and C's is going to support the PP in Madrid, why don't PSOE and PP support C's in the city of Valencia?*	T
JRL	Felicidades al @FCBarcelona Ha sido **el mejor** y nos ha hecho disfrutar. Tripletemerecido. *Congratulations to @FCBarcelona It has been the best* *and it has made us enjoy. Deserved triplet.*	A

IE	Tremendo. El Ayuntamiento de Madrid pide ***un camión extra*** para ocho contenedores más con documentos destruidos http://www.eldiario.es/_176789f2 *Tremendous. The Madrid City Council requests an extra truck for eight more containers with destroyed documents http://www.eldiario.es/_176789f2*	T
JM	El dirigente de IU Ángel Pérez firmó ***13 convenios*** con Fundación Caja Madrid. *IU leader Ángel Pérez signed 13 agreements with Fundación Caja Madrid.*	T
MM	Nueva agresión racial en EEUU: un policía amenaza con su pistola ***a una adolescente negra de 14 años*** *New racial assault in the US: a police officer threatens a 14-year-old black teenager with her gun*	T
PB	La medicina es ***un ser humano poniéndose en la piel de otro ser humano***. *Medicine is a human being putting itself in the skin of another human being.*	A
AP	Ana Palacio cree q el éxito de Podemos y Ada Colau es ***fruto d la "nostalgia por el Califato Islámico*** *Ana Palacio believes that the success of Podemos and Ada Colau is the result of "nostalgia for the Islamic Caliphate*	A
SG	El guión ya está ***escrito.*** *The script is already written.*	A
EP	El parlamento andaluz convoca ***pleno Investidura*** @_susanadiazxa el jueves a las 18h *The Andalusian Parliament convenes full Investiture @_susanadiazxa on Thursday at 6pm*	T
JO	Dep # Zerolo Buen viaje, amigo. Recordaremos ***la fuerza de tus convicciones***. *Dep # ZeroloHave a goodtrip, friend. We will remember the strength of your convictions.*	T

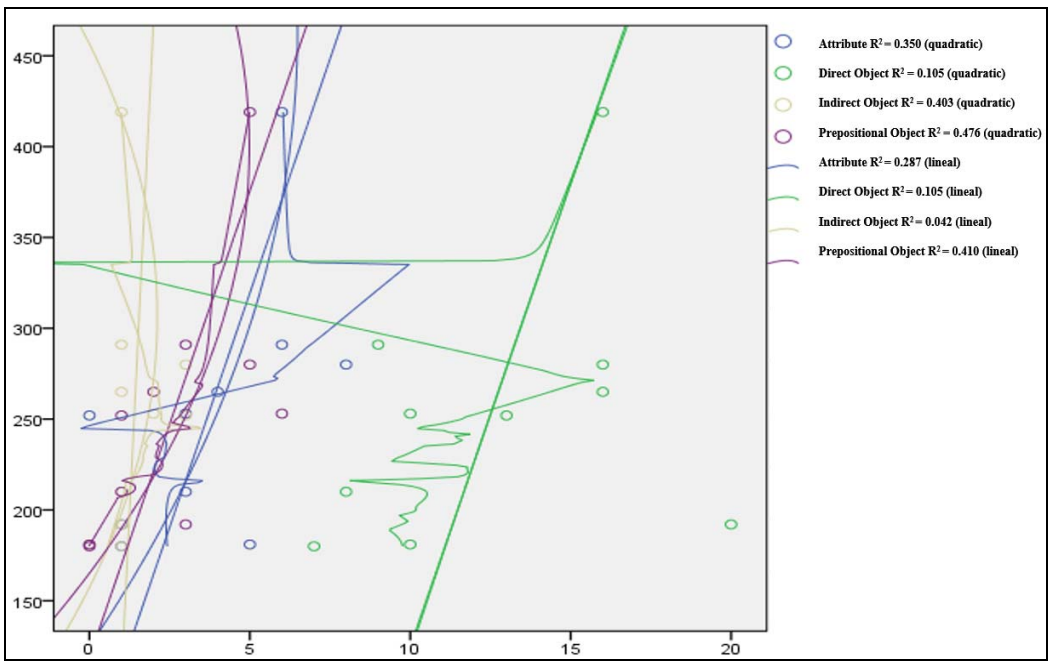

Fig. 3. Distribution of the syntactic functions in the tweets.

Table 7 presents the statistics relating to syntactic functions.

Table 7

Descriptive statistics of the density of the syntactic functions.

Statistics		*Attribute*	*Direct Object*	*Indirect Object*	*Prepositional Object*
N	Valid	10	10	10	10
	Lost	0	0	0	0
Mean		3.70	12.50	1.40	2.60
Median		3.50	11.50	1.00	2.50
Standard deviation		2.584	4.327	.966	2.171
Variance		6.678	18.722	.933	4.711
Asymmetry		.124	.375	.813	.319
Kurtosis		–.938	–1.122	–.022	–1.343
Minimum		8	13	3	6
Maximum		0	7	0	0
Total		8	20	3	6
Mean		37	125	14	26

The statistics show the frequency of the functions in relation to the total sample of the tweets (Direct Object = 12.50% / Attribute = 3.70% / Prepositional Object = 2.60% / Indirect Object = 1.40%). If we analyse the significance between the total of codified functions, we observe that the most prominent function is the direct object (n = 125 / 61.88%); and although much less prominent, the attribute is the second most widely used function in argumentative discourse structures (n = 37 / 3.70%). The kurtosis of the direct object, attribute and prepositional object is high and negative, which is in line with a platykurtic distribution, in other words, concentration around the central values of the distribution of functions in the tweets is less. This means that the tweets posted by the journalists correspond more to a syntactic structure that directly benefits the transmission of concepts rather than the valuation of such concepts. These data confirm that the syntactic structure of the tweet shares general features of standard written Spanish. A language fundamentally inclined to the active and divalent patterns which represents 57% of the total clauses of the analyzed corpus (Syntactic Database of the current Spanish, 2001). Therefore, it is confirmed that the direct object is the most common syntactic function both in general written Spanish with a percentage of 39.06% (Rojo, 2003) and in tweets (12.50%).

Linguistic structures based on the syntactic pattern is one of the most complex aspects of the computational analysis. The type of syntactic structure in each tweet is adopted in accordance with the journalist's communicative aim. For example, express opinions and feelings are normally signalled by structures with noun and adjective subordinate clauses (Noun/adjective that-clause), while those that include opinions or oppositions considerations are constructed with coordinated and subordinate clauses (for example, adversative and concessive relations) (Rudolph, 1996). We used grammatical tagging to make an initial purge of structures to obtain the following structures: noun subordinate clauses, adjective subordinate clauses, adverbial subordinate clauses, and coordinated and impersonal clauses. Figure 4 shows the density map of the syntactic structures in the tweets.

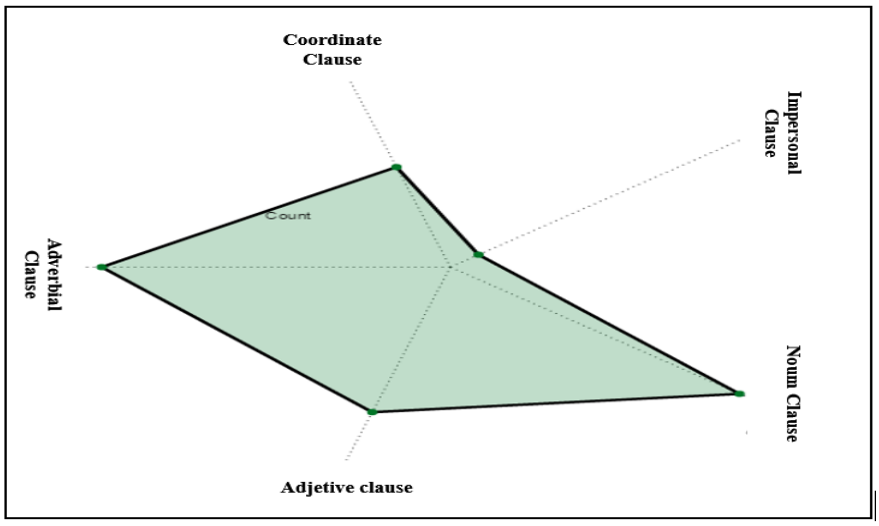

Fig. 4. *Density of the syntactic structures in the tweets.*

We observe how two structures – in particular – are used more than others: the subordinate substantive clause (n = 41 / 30.59%) and the subordinate adverbial clause (n = 40 / 30.53%). So, there is a higher percentage of these linguistic structures embedded within the overall heterogeneity of the tweet's syntactic construction. Although the journalistic tweet has no defined syntactic structure, the sample we have analysed reveals two types of recurring syntactic patterns.
Pattern 1. Noun Subordinate Clause. Example from "Jesús Maraña":
El gobernador del Banco de España cree que subir sueldos destruye empleo.
"The Governor of the Bank of Spain believes that raising wages destroys employment."
Pattern 2. Adverbial Subordinate Clause. Example from "Esther Palomera":
Ya si eso cuando sanchezcastejon *salga de Moncloa le explique alguien de qué va esto de la transparencia*
"Now if that when sanchezcastejon leaves Moncloa someone explains to you what this transparency is about."

The pattern 1 and 2 corresponding to noun and adverbial subordinate clauses allow to generate tweets like those transcribed in Table 8.

Table 8

Tweets in noun and adverbial subordinate clauses.

Journalist	*Tweets*	*N/A*[1]
PJR	Rotunda y clara @cayetanaAT en @esRadio al pedir **que se vaya el Estafermo**. Dice en voz alta **lo que la mayoría de cuadros del PP susurran**. *Strong and clear @cayetanaAT in @esRadio when requesting that the Estafermo leave. He saysoutloudwhatmost PP whisper.*	N
JRL	**Si en un breve intervalo insultan a compañeros llamándoles fascistas o acusando a tu medio de secta izquierdista**, es que vamos bien ;) *If in a short interval they insult comrades calling them fascists or accusing your left-wing sect, it is that we are doing well;)*	A

IE	El presidente del Gobierno no parece entender **en qué consiste la libertad de prensa**. Y luegohablan de Venezuela *The Prime Minister does not seem to understand what press freedom consists of. And thentheytalkabout Venezuela*	N
JM	El gobernador del Banco de España cree **que subir sueldos destruye empleo**. (Exceptosi se trata de supropiosueldo). *The Governor of the Bank of Spain believes that raising wages destroys employment. (Except if it is your own salary).*	N
PB	"**Si alguien cuenta sólo éxitos,** o no ha hecho nunca nada o miente" *"If someone counts only successes, or has never done anything or lies"*	A
AP	Ana Palacio cree **q el éxito de Podemos y Ada Colau es fruto d la "nostalgia por el Califato Islámico** *Ana Palacio believes that the success of Podemos and Ada Colau is the result of "nostalgia for the Islamic Caliphate*	N
SG	Sí...**si prometes transparencia y críticas que se negocie en los "reservados de los restaurantes**". *Ana Palacio believes that the success of Podemos and Ada Colau is the result of "nostalgia for the Islamic Caliphate*	A
EP	Pero Rajoy quiere **mejorar la comunicación** *But Rajoy wants to improve communication*	N

These patterns confirm what has been indicated previously: current Spanish clearly shows the predominance of active and divalent structures, which represent a percentage close to 60%. A substantial variation is found in the written tweet, with a more pronounced use of adverbial subordinate clauses (30.53%) with respect to the Spanish written language documented (4.24%) (Rojo, 2003). Figures 5 and 6 show the analysis of both syntactic patterns.

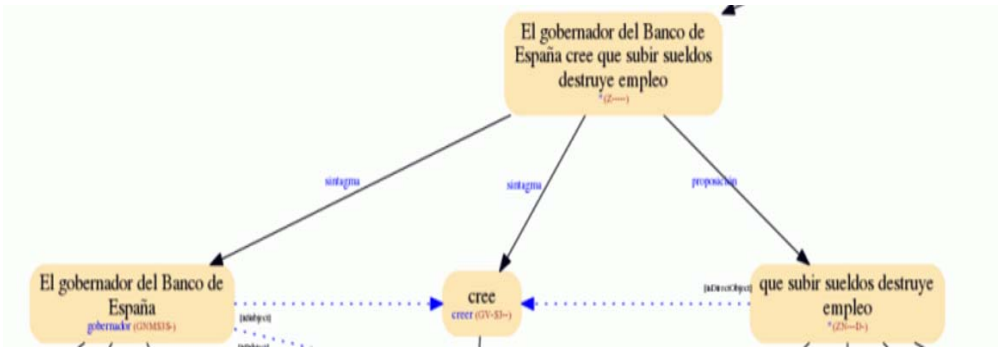

Fig. 5. Pattern 1. Noun Subordinate Clause.

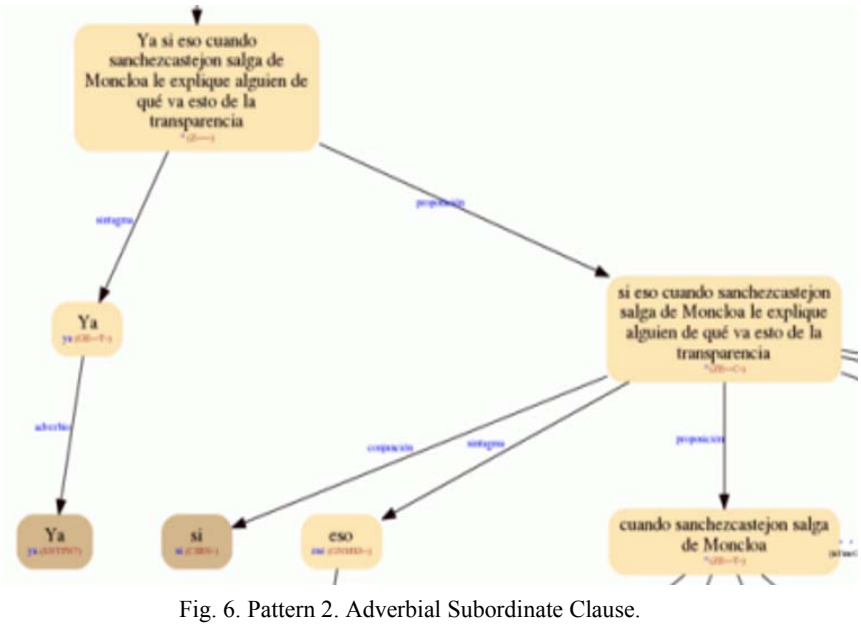

Fig. 6. Pattern 2. Adverbial Subordinate Clause.

Finally, another aspect we considered important was a demarcation to determine if there existed significant differences in the construction of the tweet according to the gender variable. To do this, as data are non-normal, we carried out a Mann-Whitney U non-parametric test; which allow us to test differences in the "distributions or differences in the "medians" of the use of word classes, the syntactic functions and the syntactic patterns of the two groups of journalists (men and women). (Tables 9, 10 y 11).

Table 9

Mann-Whitney U test (word classes).

	Noun	*Verb*	*Adjetive*	*Pronoun*	*Conjunctio*	*Preposition*	*Adverb*	*Article*
Mann-Whitney U test	9.000	11.000	7.500	5.500	7.500	8.000	9.500	10.000
Wilcoxon signed rank test	24.000	26.000	22.500	20.500	22.500	23.000	24.500	25.000
Z	–.733	–.317	–1.048	–1.467	–1.051	–.943	–.649	–.522
Asymptotic Significance (bilateral)	.463	.751	.295	.142	.293	.346	.517	.602
Exact Significance (unilateral significance)]	.548[a]	.841[a]	.310[a]	.151[a]	.310[a]	.421[a]	.548[a]	.690[a]

Table 10

Mann-Whitney U test (syntactic functions).

	Attribute	*Direct Object*	*Indirect Object*	*Propositional Object*
Mann-Whitney U test	12.500	7.500	9.000	6.500
Wilcoxon signed rank test	27.500	22.500	24.000	21.500
Z	.000	–1.061	–.827	–1.269
Asymptotic Significance (bilateral)	1.000	.289	.408	.205
Exact Significance [2* (unilateral significance)]	1.000[a]	.310[a]	.548[a]	.222[a]

Table 11

U de Mann-Whitney (sentence patterns)

	Noun C.	*Adjetive C.*	*Adverb C.*	*Impersonal C.*	*Coordinate C.*
Mann-Whitney U test	6.500	4.500	9.500	7.500	9.500
Wilcoxon signe drank test	21.500	19.500	24.500	22.500	24.500
Z	–1.261	–1.702	–.632	–1.225	–.671
Asymptotic Significance (bilateral)	.207	.089	.527	.221	.502
Exact Significance [2* (unilateral significance)]	.222[a]	.095[a]	.548[a]	.310[a]	.548[a]

The results show that the gender variable made no significant difference to the use of word classes, functions and syntactic structures since the asymptotic significance levels are not relevant up to an alpha level of 0.05.

CONCLUSIONS

Our research shows that the 10 most widely read Spanish journalists on Twitter posted an average of 20.7 tweets per day, with a formal structure of 252.3 words and 1271 characters evenly spread between men and women. The average number of words in the journalistic tweets analysedis similar to those found in generalist tweets (Hu, Talamadupula and Kambhampati, 2013), but a higher frequency than in the old SMS (Ling, & Baron, 2007).

The pre-eminence of the noun indicates that the information transmitted in the tweet is mainly conceptual. In contrast, the kurtosis is negative for the "pronoun" and "adverb", which generates a platykurtic distribution, or, a looser concentration around the central values of the distribution, meaning that these word classes are used to a lesser extent in the journalists' tweets.

We applied formal construction variables to analyse the influence and significance of word classes, syntactic functions and linguistic structures in the assembling of the journalistic tweet. The analysis shows that there are five significant word classes used by the journalists when they construct a tweet: the preposition (.824), noun (.772), verb (.705), conjunction (.678) and article (.676). The results showed that the noun is the word most commonly used (n = 716/ 31.52%), followed by the preposition (n = 410 / 18.05%) and verb (n = 336 / 14.79%). The results of the multiple linear regression test show that only two word classes recur, "preposition" and "verb", which demonstrates that regardless of the length of the tweet, these two word classes are nearly always present in the tweets. The use of the nouns in tweets coincides with its use in standard written Spanish according to the Spanish Corpus of the XXIst Century (RAE, 2015). The variation occurs in the use of the preposition and the verb. In tweets, the use of the preposition is more frequent than the verb. In standard written Spanish the verb is more frequent (Normalized frequency: 49,693.03 cases per million) than the preposition (Normalized frequency: 38,646.97 cases per million).

The analysis of the syntactic functions shows how the direct object is pre-eminent (n = 125 / 61.88%) followed at a distance by the attribute in argumentative discourse structures (n = 37 / 3.70%), as the most widely used functions. These data confirm that direct object is the most common syntactic function both in general written Spanish with a percentage of 39.06% (Rojo, 2003) and in tweets (12.50%). We also observe that

journalists tend to use two linguistic structures above all others: the noun subordinate clause (n = 41 / 30.59%) and the adverbial subordinate clause (n = 40 / 30.53%). These syntactic structures ease the transmission of concepts and the clarification of these concepts mainly with regard to causal, final and modal aspects. These patterns confirm that current Spanish clearly shows the predominance of active and divalent structures, which represent a percentage close to 60%. A substantial variation is found in the written tweet, with a more pronounced use of adverbial subordinate clauses (30.53%) with respect to the Spanish written language documented (4.24%) (Syntactic Database of the current Spanish, 2001; Rojo, 2003).

REFERENCES

Ahmad, A., 2010, "Is Twitter a Useful Tool for Journalists?", *Journal of Media Practice*, 11, 2, 145–155.

Bessho, F.,T. Harada, Y. Kuniyoshi,2012, "Dialog system using real-time crowdsourcing and Twitter large-scale corpus", *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (SIGDIAL) Seoul, South Korea, Association for Computational Linguistics, 227–231.

Bodomo, A., 2009, *Computer-Mediated Communication for Linguistics and Literacy: Technology and Natural Language Education*, Hershey, PA, Information Science Reference.

Boyd, D., S. Golder, G. Lotan, 2010, "Tweet tweet retweet: Conversational aspects of retweeting on Twitter", *Proceedings of the 43rd Hawaii International Conference on System Sciences*, Koloa, Kauai, HI, USA, 1–10.

Broersma, M., T. Graham, 2013, "Twitter as a News Source", *Journalism Practice*, 7, 4, 446–464.

Carrera-Álvarez, P., C. Saíz de Baranda, E. Herrero, N. Limón, 2012, "Journalism and Social Media: How Spanish Journalists are Using Twitter", *Estudiossobre el Mensaje Periodístico,* 18, 1, 31–53.

Carrión, M., 2013, "Un nuevo lenguaje para los medios periodísticos digitales. La necesidad de adaptarse al continuum informativo. Primeras experiencias en España", *Revista de Comunicación de la SEECI*, 32, 127–140.

Cortés-Rodríguez, L., 2012,"Los límites del discurso: condicionantes y realizaciones", *Círculo de Lingüística aplicada a la Comunicación*, 51, 3–49.

Crystal, D., 2008,*Txtng: the Gr8 Db8*, Oxford, Oxford University Press.

Dresner, E., S. Herring, 2012, "Emoticons and illocutionary force", in: D. Riesenfe, G. Scarafile (eds.), *Philosophical dialogue: Writings in honor of Marcelo Dascal*, London, College Publication, 59–70.

Fletcher, W. H., 2004, "Facilitating the compilation and dissemination of *ad hoc* web corpora", in: G. Aston, S. Bernardini, D. Stewart (eds), *Corpora and Language Learners*, Amsterdam, John Benjamins, 273–300.

Fortunati, L, 2001, June, *The mobile phone between orality and writing,* paper presented at the *Third International Conference on Uses and Services in Telecommunications*, Paris, France.

Fraca de Barrera, L., 2006, *La ciberlingua. Una variedad compleja de lengua en Internet*, Caracas, Instituto Venezolano de Investigaciones Lingüísticas y Literarias "Andrés Bello".

Gómez-Camacho, A., 2007, "La ortografía del español y los géneros electrónicos", *Comunicar*, 29, 157–164.

Gómez-Camacho, A., 2014, "La norma disortográfica en la escritura digital", *Didac*, 63, 19–25.

Hernando, L., 1994, "Comunicación y lenguaje en el periodismo escrito", *Didàctica,* 6, 145–159.

Honeycutt, C., S. C. Herring, 2009, "Beyond microblogging: Conversation and collaboration via Twitter", *Proceedings of the 42nd Hawaii International Conference on System Sciences*, Los Alamitos, CA, IEEE Press, 1–10.

Hong, L., G. Convertino, E. Chi, 2011, "Language matters in twitter: A large scale study", *International AAAI Conference on Weblogs and Social Media*, 518–521.

Horstmanshof, L., M. R. Power, 2005, "Mobile phones, SMS, and relationships", *Australian Journal of Communication*, 32, 1, 33–52.

Hu, Y., K. Talamadupula, S. Kambhampati, 2013, "Dude, srsly?: The Surprisingly Formal Nature of Twitter's Language", *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media* Cambridge, Massachusetts, USA, AAAI press, 244–253.

Hutchby, I., V. Tanna, 2008, "Aspects of sequential organization in text message exchange", *Discourse and Communication*, 2, 2, 143–164.

Inaba, M., S. Kamizono, K. Takahashi, 2013, "Utterance generation for non-task-oriented dialogue systems using Twitter", *Proceedings of the 27th Annual Conference of the Japanese Society for Artificial Intelligence*, 1K4-OS-17b-4, 334–338.

Jaffe, A., S. Walton, 2000, "The voices people read: orthography and the representation of non-standard speech", *Journal of Sociolinguistics*, 4, 4, 561–587.

Java, A., X., Song, T., Finin, B. Tseng, 2007, "Why We Twitter: An Analysis of a Microblogging Community", in: H. Zhang *et al.* (eds), *Advances in Web Mining and Web Usage Analysis. Lecture Notes in Computer Science*, Berlin, Springer, 118–138.

Lasorsa, D., S. C. Lewis, A. Holton, 2012, "Normalizing Twitter: Journalism Practice in an Emerging Communication Space", *Journalism Studies*, 13, 1, 9–36.

Lázaro Carreter, F., 1977,"El lenguaje periodístico entre el literario, administrativo y vulgar", in: F. Lázaro, (ed.), *El lenguaje en el periodismo escrito,* Madrid, Fundación Juan March, 9–32.

Lewis, C., B. Fabos, 2005, "Instant messaging, literacies, and social identities", *Reading Research Quarterly*, 40,4, 470–500.

Ling, R., N. Baron, 2007, "Text Messaging and IM: A Linguistic Comparison of American College Data", *Journal of Language and Social Psychology*, 26,3, 291–298.

Lomborg, S., 2011, *Negotiating the Twitter self. On networks of affiliation and relational pressures*, Boston, MA, International Communication Association.

Mancera-Rueda, A., A. Pano-Alamán, 2013, *El español coloquial en las redes sociales*, Madrid, Arco/libros.

Markman, K. M., 2013, "Conversational coherence in small group chat", in: S. Herring, D. Stein, T. Virtanen(eds), *Pragmatics of Computer-Mediated Communication*, Berlin, Mouton de Gruyter, 539–564.

Menna, L., 2012, "Nuevas formas de significación en red: el uso de las #etiquetas en el movimiento 15M", *Estudios de Lingüística del Español*, 34, 1–61.

Pano-Alamán, A., A. Mancera-Rueda, 2014, "La conversación en Twitter: las unidades discursivas y el uso de marcadores interactivos en los intercambios con parlamentarios españoles en esta red social", *Estudios de Lingüística del Español*, 35, 1, 234–268.

Parodi, G., 2010, *Lingüística de corpus: de la teoría a la empiria*, Frankfurt, Iberoamericana/Vervuert.

Real Academia Española (2015). *Corpus del Español del Siglo XXI (CORPES)*. Retrieved from http://web.frl.es/CORPES/view/inicioExterno.view

Riordan, M., K., Markman, C. O. Stewart, 2013, "Communication accommodation in instant messaging: An examination of temporal convergence", *Journal of Language and Social Psychology*, 32, 1, 84–95.

Rojo, G., 2003, "La frecuencia de los esquemas sintácticos clausales en español", in: F. Moreno Fernández; F. Gimeno Menéndez; J. A. Samper; M.ª L. Gutiérrez Araus; M. Vaquero, C. Hernández (Eds.), *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*, Madrid, Arco/Libros, 413–424.

Rudolph, E., 1996, "Adversative and Concessive Relations and their Expressions in English, German, Spanish, Portuguese on Sentence and Text Level", Berlin, De Gruyter. https://doi.org/10.1515/9783110815856

Sugiyama, H., T., Meguro, R., Higashinaka, Y. Minami, 2013, "Open-domain utterance generation for conversational dialogue systems using web-scale dependency structures", *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (SIGDIAL) Seoul, South Korea: Association for Computational Linguistics, 334–338.

Syntactic Database of the current Spanish (2001). *Base de Datos Sintácticos del español actual (BDS)*. Retrieved from http://www.bds.usc.es/

Thurlow, C., K. Mroczek, 2011, *Digital discourse: Language in the new media*, New York, Oxford University Press.

Thurlow, C., M. Poff, 2011, "Text-messaging", in: S. Herring, D. Stein, T. Virtanen (eds), *Handbook of the pragmatics of CMC*, Berlin and New York, Mouton de Gruyter, 1–24.

Vázquez-Cano, E., 2012, "Mobile Learning with Twitter to Improve Linguistic Competence at Secondary Schools", *The New Educational Review*, 29, 3, 134–147.

Vázquez-Cano, E., J., Fombona, C. Bernal, 2016, "Análisis computacional de las características ortipográficas y paralingüísticas de los tweets periodísticos", *El profesional de la Información*, 25, 4, 588–598.

Vázquez-Cano, E., E., López Meneses, M.ª L. Sevillano,2017,"La repercusión del movimiento MOOC en las redes sociales. Un estudio computacional y estadístico en Twitter", *Revista Española de Pedagogía*, 75, 266, 47–64.

Vázquez-Cano, E., S., Mengual-Andrés, R. Roig-Vila, 2015, "Análisis lexicométrico de la especificidad de la escritura digital del adolescente en Whastapp", *Revista de Lingüística Teórica y Aplicada*, 53, 1, 83–105.

Wang, A., Chen, T. Kan, M-Y, 2016, "Re-tweeting from a linguistic perspective", *Proceedings of the Second Workshop on Language in Social Media*, Montréal, Canada, 46–55.

Yoshino, K., S., Mori, T. Kawahara, 2011, "Spoken dialogue system based on information extraction using similarity of predicate argument structures", *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (SIGDIAL) Seoul, South Korea, Association for Computational Linguistics, 59–66.

Yus, F., 2001, *Ciberpragmática. El uso del lenguaje en Internet*, Barcelona, Ariel.