

AN INSIGHT INTO TWITTER: A CORPUS BASED CONTRASTIVE STUDY IN ENGLISH AND SPANISH

Irina Argüelles Álvarez
Alfonso Muñoz Muñoz

Universidad Politécnica de Madrid

Abstract: *The aim of this paper is to study the use of Spanish and English in the micro-blogging social network Twitter from a contrastive point of view. A quantitative research methodology is applied in order firstly, to identify specific common characteristics of language, organization and content in the medium and secondly, to find eventual differences in the use of a particular language. To carry out the experiment, two corpora were constructed using language data from Twitter, one in Spanish with a total number of 4,027,746 words and another with similar characteristics in English with a total number of 4,655,992 words. From the results obtained, the conclusion is that there are a number of very general discourse and organizational features common to the two corpora under study. It is also concluded that there are some particular characteristics which differentiate the use of English and Spanish in the medium.*

Key words: *Twitter, contrastive analysis, corpus linguistics.*

1. INTRODUCTION

The Web 2.0 concept was first presented in 2004 by O'Reilly to refer to a second generation web based on communities of users and a special range of services such as social networks, blogs or wikis that promote collaboration and the exchange of information among users in an easy and rapid way. A social network is seen by Boyd and Ellison (2007) as a well defined set of actors, individuals, groups, organizations, societies, etc., that are linked by a number of social relations, simplified in the case of IT by the use of new communications technologies. Other studies reflect a social reality centered on the relations between individuals. A social network can be described based on its efficiency to integrate three concepts (the 3Cs): Communication (to share knowledge and make it public), community (to find and to link communities) and cooperation (to do things together) (Misanchuk and Anderson 2001). According to Levy, Little and Aiyegbayo (2007), a social network is the use of Internet applications to connect people in communities of interest and resource exchange. This concept of "community of interest" makes the linguist think of a potential discourse community basically defined by the medium used to communicate, the communicative needs and the motivation of people to join the community (Swales 1990). Although social networks imply the idea of global exchange of information, most network applications are in fact limited to a group of users who basically share similar work or leisure interests.

The concepts of discourse community and communicative purpose as characterized by Swales (1990: 45-57) as well as a number of discourse conventions (Bazerman, 2000) are here adopted to locate the genre by means of the most immediate parameters: the sender, the receiver, the channel and the purpose of communication.

A genre comprises a class of communicative events, the members of which share some set of communicative purposes. [...] Communicative purpose is both a privileged criterion and one that operates to keep the scope of a genre as here conceived narrowly focused on comparable rhetorical action. [...] If all high probability expectations are realized, the exemplar will be viewed as prototypical by the parent discourse community. (Swales 1990: 58)

Twitter has a number of specific characteristics that makes it different from others the most obvious being probably that the network allows the user to send “tweets” with a maximum length of 140 characters. But apart from the length of the messages, Twitter has also established among its users a well-defined markup culture: RT stands for “retweet”, a mechanism that empowers users to spread information of their choice beyond the reach of the original tweet’s followers; #followed by a word represents a “hashtag” which categorizes relevant keywords in a tweet to show more easily in twitter search and @followed by a user identifier address the user, a marker of addressivity which facilitate exchanges. With respect to this last feature, it is interesting to notice that the relationship among users in Twitter, unlike other social networking sites such as Facebook or Tuenti in Spain, does not require reciprocation which means that one user can “follow” any other user, and that “followed” user need not follow back. Every Twitter user can have a certain number of “followers”, people who read the author of a tweet, and the number of followers a person has, defines, to a certain extent, his or her popularity in the network (O’ Reilly and Milstein 2009). Reading messages other people have written is known as “following”. According to Kwak et al. (2010: 593) 77.9% of user pairs with any link between them are connected one-way and only 22.1% have reciprocal relationship whereas previous studies have reported much higher reciprocity on other social networking services. These results lead to the conclusion that Twitter is to a certain extent, more a source of information than a social networking site.

The study presented here introduces an analysis of the language used in a specific micro-blogging social network site which is about five years old, commands more than 100 million active users in October 2011, and is growing fast. This social network site has been selected because of its specific characteristics based on brief communications, which differ from other social network sites such as Myspace, Facebook, Orkut, Friendster or Bebo. In the last years, Twitter has become one of the most popular media to provide real time information about the most outstanding happenings world-wide; nevertheless, Twitter users mostly “tweet” (the action to post text based messages) about other much more routinary happenings in their lives. In order to identify global characteristics which describe Twitter and to find structural, discursive or pragmatic features which distinguish the local use in English and Spanish, a number of measurements were carried out in two parallel corpora where about four million words in English and Spanish were tagged. Six hundred thousand sentences and approximately 46 million characters were included in total. Conclusions from the study of this collection of texts will be presented from a contrastive point of view.

2. METHODOLOGY OF THE STUDY

The twofold objective of this study is to describe Twitter from a discourse contrastive point of view and to study a number of distinguishing features of language use that are specific to this social network in Spanish and English. To reach these objectives, an analysis methodology which includes linguistic hypotheses, measurements of frequency in real corpora and manual analysis of specific examples, has been carried out.

2.1. The two parallel corpora

The construction of the corpora is based on the messages published by 206 users (103 in Spanish and 103 in English). The messages were compiled by means of a crawler programmed in Java. This program automatically converted the web page format into plain text (.txt files), facilitating analysis at a later stage. The selection of messages was established firstly to be from users with a minimum number of 1,000 followers as individuals with around this number of followers are the ‘most popular’ Twitter users. Secondly, users with approximately 3,200 tweets published. Although it is in fact possible to go on adding tweets to the 3,200 published, once

the user has arrived at this figure, the initial entries are automatically erased in order to permit the publication of the latest ones. Hence, popularity and number of tweets were the criteria established to determine a set of messages that a large number of users want to read and which use standard language and not the slang of a limited group.

The subject matter and the homogeneity of the corpora were difficult to establish because, as it is by now mostly known, a Twitter user can write about any subject in each tweet, which may be unrelated to the previous message. Despite this, it was identified that there were messages about technology, which were common to both languages, and therefore, users with a technological profile were selected. Consequently, the corpus created in Spanish consists of 319,381 Twitter messages and a total of 4,027,746 words. In English there are 330,696 Twitter messages and a total number of 4,655,992 words. In order to get conclusions from the use of different linguistic categories, the corpora were tagged with a program implemented in JAVA language that uses the tagger *TreeTagger* (Schmidt 2009), developed at the Institute for Computational Linguistics of the University of Stuttgart. This tagger was selected because of its accuracy and free access permissions. Although for some linguistic measurements in the corpora the limitations or imprecision of this tagger must be taken into account, it is considered to be reliable enough for an initial analysis of the quantitative data.

3. QUANTITATIVE ANALYSIS OF DATA

A frequent user of the micro-blogging network Twitter could no doubt list a number of very general typical characteristics of the medium, from the simple observation with the naked eye: it consists of very brief communications which make use of web links to provide further information about the where and when of different events, the nature of the relation among users is based on followings and followers and there are specific means of exchanging messages. Some of these salient features are made explicit by a number of markups such as @ (at), # (hashtag) or RT (retweet), and the maximum length of 140 characters actually limit the communications.

These characteristics, typically format characteristics of tweets, can in fact be quantified to help establish discursive hypotheses about the use of the language in the medium. The different data obtained from the quantitative approximation and the comparison of the results between languages, actually provided initial assumptions for a subsequent more in depth qualitative study of different specific linguistic and discursive aspects presented in what follows in different subsections together with examples from the corpus.

In order to fulfill a preliminary objective, to give evidence with exact data of the basic structural characteristics of tweets, measurements related to their typical format were done. The data obtained from this analysis, such as the average length of the sentences, words per sentence or characters per word are presented in Table 1. The aim was to set an initial quantitative approximation to describe the corpus and to observe very general regularities or differences in length. These measurements of sentence and word length were of potential interest to the study because, as it will be subsequently explained, the restriction of the medium in the number of characters per sentence seemed to have an effect in the language used.

Table 1. Average values of words and characters per sentence in the corpora.

	Corpus in Spanish	Corpus in English
Number of users	103	103
Tweets downloaded	319,381	330,696
Number of words	4,027,746	4,655,992
Number of sentences	319,381	330,696
Number of characters	21,519,161	25,010,795
Number of words/sentence	12.61	14.07
Number of characters/Word	5.34	5.37
Number of characters/sentence	67.37	75.63

With regard to the number of words and the length of the sentences, the first analysis indicated that, on average, users of Twitter make use of only a 51.07% of the total maximum capacity allowed (67.37 characters in Spanish and 75.63 in English are used from the 140 characters maximum per sentence allowed).

Although tweets of the maximum length are scant, here are some examples of messages that can give a general idea of the amount of information and structure that a tweet of this length has:

1. @luiscesarmaza exacto, Twitter es la demostración de que cualquier pequeño microcambio puede generar grandes cambios... <http://bit.ly/NtSgm>
2. RT @jayrosen_nyu: what I said to MSNBC: "Twitter now is a more effective system than any single news organization at serving breaking news."

See the following as examples of average length sentences:

3. "Twitter for me is a news broadcast channel" @guykawasaki #openwebcamp
4. twitter ahora vale mil millones de dólares <http://bit.ly/1Q7pC0>

After compiling pure structural information, measuring linguistic categories with *Tree Tagger* (see Table 2) was also found useful in order to count with data from where to confirm initial expectations or general impressions. Then, these measurements were complemented with a count of symbols and some specific typical Twitter terminology which would help to determine the resources that confer a permanent sensation of immediacy to the medium (Tables 3 and 4).

Table 2. Result of the quantitative count of different linguistic categories according to the tags provided by the software used, *TreeTagger*.

Corpus in Spanish			Corpus in English		
Total number of words: 4,027,746			Total number of words: 4,655,992		
Nouns	1,284,509	31.89%	Nouns	1,888,343	40.55%
NC	863,141	21.42%	NC	1,126,624	24.19%
NP	417,859	10.37%	NP	761,719	16.35%
Adjectives	246,313	6.11%	Adjectives	322,857	6.93%
Adverbs	142,192	3.53%	Adverbs	242,902	5.21%
place	12,411	0.30%	place	11,624	0.24%
time	36,713	0.91%	time	35,839	0.76%
manner	10,747	0.26%	manner	19,001	0.40%
quantity	42,687	1.05%	quantity	45,688	0.98%
Verbs	800,093	19.86%	Verbs	1,003,366	21.54%
infinitives	137,575		infinitives	170,572	
gerunds	12,078		gerunds	16,053	
participles	136,249		participles	169,568	

Table 3 below, presents the total results for different typical and frequent characters or symbols described before as habitual in Twitter. Although the departing idea is that the users of the medium will make use of these symbols or characters independently of the language, the number of times they are in fact used can lead to conclusions regarding local differences.

Table 3. Total number of special characters and symbols and percentage with respect to the total number of words

Corpus in Spanish			Corpus in English		
Emoticons ¹	24,370	0.61%	Emoticons	17,428	0.37%
http addresses	94,600	2.34%	http addresses	104,729	2.24%
@+Word	200,516	4.98%	@+Word	289,114	6.20%
RT	28,889	0.72%	RT	51,770	1.11%
Word starting#	3,008	0.07%	Wordsatrtng#	3,814	0.08%

Finally, as the symbols or special characters in Twitter mostly appear either starting or closing tweets, their position in the sentence is also studied and presented in Table 4.

¹ 125 emoticons were counted. Some examples from the list are the following: :-) (-: :) (= =:D =D xD :-(-): :(): D: Dx :-P =P :(='[:_(=O :-O :-X :-;) :-D 8-) :-o :O :S xD :< 3<, etc

Table 4. Sentence beginnings and endings in Twitter.

	Corpus in Spanish		Corpus in English	
Total number of sentences	319,381		330,696	
Sentences starting with:				
capital letter	148,021	46.34%	167,334	50.60%
@	123,186	38.57%	127,756	38.63%
Number	2,356	0.73%	3,840	1.16%
RT ²	20,205	6.32%	39,137	11.83%
lower case letter	32,610	10.21%	20,789	6.28%
#	3,008	0.94%	3,814	1.15%
Sentences finishing with:				
full stop	58,297	18.25%	74,860	22.63%
Dots	29,596	9.26%	6,157	1.86%
Emoticon	19,841	6.21%	13,547	4.09%
http address	74,204	23.23%	69,733	21.08%

In table 4 (beginning and endings of tweets), some of the typical abbreviations which help the organization of discourse with a maximum economy of words have been included. The situation of these symbols or markups at the beginning or the end of the tweets is regular across the examples in the corpus. From the data, it is understood that the frequency of use of the symbols and abbreviations, with very few exceptions, is very similar in both languages.

4. FINDINGS

From the data presented in the previous subsection, useful conclusions can be drawn. The fact that the length of the sentences does not exhaust the total maximum capacity allowed by the medium confirms the first impression of its immediacy and its ability to communicate very specific actions, news or events that take place at a specific moment in time. From the data, it is also confirmed the general impression that the tweets in English are usually longer than in Spanish, English uses more words per tweet. Therefore, the next step will be to arrive to conclusions about the nature of these words and their function in the tweets.

From the data obtained in this quantitative initial measuring, the following specific linguistic and discursive characteristics will be further analyzed:

- The most frequent category in both languages, nouns (31.89% in Spanish < 40.55% in English), can provide the study with the main themes people tweet about.
- The use of verbs is also high and together with nouns, they increase the number of total words in English when compared to the total in Spanish. The percentage of gerunds is low (Table 2) although the initial hypothesis was that these would appear in a higher percentage as the tendency was supposed for Twitter users to write about what one *is doing* at the time of writing or around that time.

² Mesarements for RT are included also in the resulting for "Capital letters".

- Adverbs are also used in English in a higher percentage (5.21%) compared to Spanish (3.53%), the initial hypothesis from the quantitative data being that the use of adverbs has an inverse relation with the use of emoticons and other marks such as dots: Spanish uses more emoticons together with other marks and less adverbs than English (4.09% emoticons in English < 6.21% in Spanish and 1.86% dots in English < 9.26% in Spanish). On the contrary, more adverbs and fewer emoticons and other marks are used in English.

4.1. Themes users tweet about

Starting with the most frequent grammatical category in the corpus, nouns related to time predominate (time, days or parts of the day). This finding redounds to the idea of immediacy in this medium. Other important themes are common to both languages and these can be grouped under the labels “routines” (common places, work or leisure) and “technology” (social networks and devices, past news, events and actuality / what is going on) both are recurrent topics in the corpora. See below some of the most frequent nouns from the corpus in English:

Time (corpus in English)> time 8576, day 7816, today 6777, morning 3549, night 3169, year 2465, days 1969, weekend 1434, hours 1359, times 1349, Friday 1156, minutes 1030>

Routines (corpus in English)<work 2679, fun 2645, business 2348, home 2284, friends 2222, party 2098, book 2070, kids 2054, news 2028, job 1505, school 1127, office 1085, movie 1000>

With respect to technology, both in English as well as in Spanish, the second most frequent word is “Twitter” (12,016 in English> 9,869 in Spanish) following “thanks” (12,156 in English>9,906 in Spanish), the first word in the ranking. Other frequent words related to technology are:

Technology (corpus in Spanish) <Twitter 9869, blog(s) 6687, google 3933, vídeo(s) 3174, internet 3023, facebook 2937, red(es) 2895, web 2183, iphone 2045, youtube 1040, Mac 1012>

Also connected both with time and the topic of leisure, it is interesting to find that the only day of the week with an occurrence frequency over 1,000 in both languages is Friday. It could be hypothesized that cultural reasons can in fact explain the different occurrence frequency of the days of the week in the corpora. According to Stubbs (2001: 16) the actual occurrence of the seven days of the week in a corpus varies due to cultural reasons. From his data, people speak more often about the weekend, less often about the beginning and end of the week and the least often about the rest of the days. The conclusions from the data obtained only partially agree with those by Stubbs.

Table 5. The days of the week presented in an occurrence frequency order.

Corpus in Spanish		Corpus in English	
Viernes (Friday)	1,013	Friday	1,156
Lunes (Monday)	848	Sunday	926
Domingo (Sunday)	631	Monday	826
Jueves (Thursday)	614	Saturday	572
Sábado (Saturday)	543	Tuesday	422
Martes (Tuesday)	468	Thursday	346
Miércoles (Wednesday)	384	Wednesday	333

It is a fact that Friday has become part of the weekend and almost the peak leisure day, or more precisely night, among youngsters and also adults:

5. You heard about da Phoenix show in Vancouver on *Friday night* right? The one where the ENTIRE audience got on stage? :P <http://bit.ly/aiOzkt>
6. ¿Quien se une al #twittmuseo? Este *viernes noche* visita a museos de Sevilla. Empezaremos por el de Bellas Artes ¿no?

But whereas Thursday has almost become part of the weekend in Spain and it occurs even more than Saturday in the Spanish corpus (614), it is the last day in frequency in the English one (346).

7. Good Music in NYC this *Thursday night* with @MiekaPauley <http://bit.ly/8hjR01>
8. Me invitan a una fiesta el *jueves por la noche* con gente interesante. YUHU!

The use of adverbs of time in the corpora also point to the importance of the temporal aspect in Twitter. Going back to Table 2, it is remarkable that adverbs of time are used with a very high frequency in both languages and the analysis of those adverbs redounds to the same general effect regarding the importance of time in the medium. Table 6 below summarizes the tokens counted for the most frequent in both corpora.

Table 6. Adverbs of time.

Now	12,122	Ahora	7,088
Today	8,107	Hoy	7,523
After	2,873	Después	2,472
Tomorrow	2,642	Mañana	2,636
Before	2,562	Antes	2,183

“Now” / “ahora” and “today” / “hoy” lead once again to conclusions about the importance of the immediate temporal context for Twitter users.

9. *Now* watching HTML5 video in IE9. The audience cheers.
10. *Ahora* viendo a Fernando Berlín (@radiocable) en VEO TV
11. @PeterRosdahl 12:30 *today* Ballroom D #sxsw
12. *Hoy* a las 19h30 con Javier Godoy en La Casa del Libro de Gran Vía, en Madrid &&&
<http://7wiep.tk>

Time has demonstrated to be the most relevant factor in Twitter and leisure, routines and technology can in fact be listed as the commonest topics in the corpora. Contrary to the expected difficulty to identify the most frequently treated themes in the medium, these can be neatly identified from the analysis of noun frequency and the data presented hereby agree with previous studies aimed at identifying individual user’s intention in using this technology (Java et al. 2007).

4.2. The use of verbs

From the original data, and regarding the use of verbs, two conclusions are to be illustrated in this subsection with data from the corpora: first, that the use of verbs is higher in English when compared to their use in Spanish; second, that the percentage of gerunds is low. This could mean that Twitter users do not write about what they *are doing* at the time of writing or around that time more than about what they have done, will do, or any other communicative purpose they may have.

Connected with the length of the sentence discussed in previous subsections, is probably the use of verbs. The total number of words (Table 1) and the number of verbs in English and Spanish (Table 2) reinforce the same idea that tweets in English are in general longer. It could be hypothesized that in this medium, Spanish is more adaptable to the omission of parts of the speech. English seems to have a greater tendency to produce grammatically correct full sentences.

13. Jeremy gave me the latest update on what's going on at @fourbarrel & in the back I noticed that they now have a new @SlayerEspresso

It has been found more difficult to locate examples of sentences with no verbs in English, examples that are quite usual in Spanish. Notwithstanding, verb omissions can be found in both languages for reasons of space economy, the last forced by the maximum number of characters per tweet:

14. En vivo en los Encuentros Digitales de El Economista: <http://j.mp/6QI2D3>
15. The scene here from inside the #nasatweetup tent <http://bit.ly/1NVcZk>
16. Algunas fotos del #bymc <http://bit.ly/gAdwq>
17. A few photos from the #nasatweetup at the Kennedy Space Center in Florida for @nasa Space Shuttle launch of STS-129 <http://bit.ly/41JmXU>

Now, turning to other aspects regarding the use of verbs, hypothesis about a possible overuse of the gerund in comparison with other tenses were not confirmed for any of the languages (Table 2). The initial hypothesis derived from the possible tendency to write about what one *is doing* at the time or around the time of writing according to the initial prompt on Twitter home page that used to urge users to answer to the question: What are you doing?

Here are some examples from the corpora of the structure expected:

18. *Probando* twitterphone. Me gusta.
19. *Testing* my htc I love it so far!
20. *Saliendo* del hotel rumbo al #BYMC:D
21. Ok, *checking out* of Ellington Hotel, heading to airport, and back to Eastern Time. Man, it's been a long trip.
22. *Descansando* en el hotel en Dublín
23. #next10 is over; back at hotel *resting* before dinner.

However, the results obtained agree with the conclusions by Mischaud (2007) that 58.5% of the 5,767 messages studied did not address this question. Probably it was in fact this lack of coherence of the actual tweets with the original question that led Twitter homepage to change the former “What are you doing?” by the latest version “Follow your interests” (twitter.com). And with “interests” the direction of thought moves back to the most frequent nouns analysed in the previous subsection and also again back to time.

4.3. Emoticons and stance

The frequent use of emoticons and other symbols (see Tables 3 and 4) can provide more information about discursive or even pragmatic aspects which characterize Twitter. From the Tables above, it is concluded that Spanish uses more emoticons and final dots in comparison with English. Emoticons make the “real” meaning of the utterance more obvious by adding to the phrase a feeling about what has been said as for example: ☺ meaning “I say it happily” - 😞, ☹ “I say it sadly” - :(, “I am joking” - ;) . The use of emoticons is related to feelings or evaluation about what is said. From a discursive point of view, some of these attitudes towards what is said are explicitly expressed in English by means of evaluation adverbs or phrases:

24. Amigos, fútbol, cerveza y palomitas ☺
25. Beer, sunshine and chocolate enough *to make anyone smile*.
26. Cogiendo setas - en esta época, esto solo pasa en Galicia ☺
27. Picking up lots of New Friends & Followers on Facebook & Twitter - *Great Event!*
28. @josempelaez pues parece que éste lo consigue... ☹
29. @netspencer Ugh, *sadly* no.

From the data in Table 2, it can be seen that English uses more adverbs and in particular adverbs of manner, which could in fact be replacing the meaning of the emoticons that are more common in Spanish according to the results in Table 3. In Tables 7 and 8 the results obtained for the number of occurrences of the adverbs “sadly” and “happily” in English and in Spanish are compared with the number of emoticons to express the same feeling.

Table 7. Emoticons expressing sadness.

Sadly	104	Tristemente	2
Sadness, crying emoticons	1,374	Tristeza (sadness), llanto (crying)	3,190
Emoticons			
:(1,096	:(2,174
:-)	165	:-)	833
6):	85	:')	53
:')	12	:_)	46
;-;	8	u_u	42
:\	8	:\	25
		u.u	17

Table 8. Emoticons expressing happiness.

Happily	62	Felizmente	5
Happiness, Smile emoticons	20,833	Felicidad (happiness), sonrisa (smile)	21,726
Emoticons			
:)	14,848	:)	15,417
:-)	4,329	:-)	5,869
=)	1,456	^^	175
_	102	x-)	138
^_^	87	^_^	110
(:	6	=)	17
^^	5		

The total number of emoticons expressing positive attitude in English and Spanish represent ten times the total number of emoticons expressing negative attitude in both languages (42,559 > 4,564). From the data, it is also concluded that Twitter users' reasons for writing their tweets are more positive than negative which would point to a general positive evaluative prosody in the discourse:

30. @Ril_Juice rofl...and good night. :)

31. Hora de irse a la cama. Buenas noches :)

In the examples above, "good night" (buenas noches) would be understood to have a neutral prosody attending to its propositional meaning but the emoticon J adds positive discourse prosody. Many more similar examples are found in the two corpora:

32. @musicbizkid I'll keep u posted! :)

33. @eliwhit Haha... you didn't say I couldn't @ reply you. :)

34. @kylelreed @austinklee Jesus is on that list. :)

35. @PAJoseph Thanks for the heads up. I didn't know! :)

36. @meredithkoonce Thank you!! :)

37. @juanlarzabal Yo soy su fan número 1000 :)

38. Nacho Vegas el martes en Hoy Empieza Todo :)

39. Imagino que esto es estar inspirado :)

40. @soniagarcia Me llevas mi pancarta? :)

41. @nazaret Para mantener las apariencias :)

With regard to cultural aspects, it is also noticeable the preference of the different languages for different symbols to express the same feeling or emotion, and the wider range of feelings expressed through emoticons in Spanish:

- 102 in English > 0 in Spanish to express happiness
=> 1456 in English > => 17 in Spanish to express happiness
↪ 222 in Spanish > 0 in English to express anger
:-o 345 in Spanish > 12 in English to express surprise

42. RT @loretahur He puesto el móvil a buscar bluetooth y he encontrado más de 30 dispositivos :-o

43. @KimmieTex I do. It doesn't. *Surprisingly.*

In addition, concerning the use of dots closing sentences, as do emoticons, the former are here understood as more than mere punctuation marks. Data analysis reveals that the difference of frequency in their use between English and Spanish positively relates to the use of emoticons. In both languages the dots could be interpreted to signal attitude towards the idea stated as, most times, their usage does not fit exactly with their strict grammatical function³.

44. Atocha. Coche y a casita...

45. Alpine, Texas. Nice to be home again...

46. Cuanto más juego con la storm 2 más me gusta...

47. Started using Google Chrome instead of Firefox. liking it...

This pragmatic use of the dots can explain why they are used more extensively in Spanish than in English, as is the case with emoticons. From the data, it could be thought that there is a paralinguistic use of symbols in a text-based medium as is Twitter, similar to the use of intonation and gesture in speech, to reinforce the expression of attitudinal stance⁴.

Other realizations of this stance category in Twitter can also be exemplified again by the use of adverbs. According to Tseronis (2009: 49) single word adverbs constitute the most varied grammatical realisation of stance in English and are more frequent than other grammatical realisations of stance: <unfortunately 178 incredibly 147 sadly 104 surprisingly 65 happily 62 amazingly 48 luckily 46 fortunately 40 strangely 37 thankfully 34 unbelievably 31 dramatically 25 wonderfully 21 romantically 20> (858)

From the examples in the corpora, it is suggested that emoticons and other symbols or paralinguistic devices are used in a different way in English and Spanish which leads to conclude that there are differences in the expression of attitudinal stance between languages. Emoticons are highly used in the medium, more in Spanish, to express stance, again, with the maximum economy, and, from the data, probably replacing eventual stance adverbs or phrases in both languages.

³ According to the Cambridge International Dictionary of English (1995: 413) "Dots are not used very often. They might be used to show that:
• The writer has omitted some of another person's words [...]
• A list or an idea has not been completed [...]
• Some time passes between a speaker's words [...]
• A speaker stopped or was interrupted [...]"

⁴ The notion of stance (Conrad and Biber 2000: 57) includes three broad categories: epistemic stance (certainty/doubt), style stance (discourse comments), and attitudinal stance (positive/negative attitudes/feelings).

5. CONCLUSIONS

There are a number of very general discourse and organizational characteristics common to the two corpora under study in Spanish and English. First, Twitter users are aware of the immediacy of the communications in the medium and the restriction of characters per tweet has led them to create a number of specific codes that permit to organize discourse with maximum economy. The most common have been presented and it has been noticed how these codes (markups) help to organize discourse pointing to parts said by someone before (RT), addressing to someone else (@e-mail address) or referring the audience to web-pages where to find more information (http addresses). These abbreviations and other symbols or sentence types that characterise this particular genre, typically open or close the tweets.

Second, time is a crucial contextual factor as demonstrated by the use of different linguistic categories either to situate the messages in a temporal context or to provide information within the message about the “when” of different actions or events.

Third, general frameworks can be recognized as demonstrated by the use of a number of words related to certain topics such as: routines, news / events and IT. Although it was expected to find tweets telling what people “are actually doing” at or around the moment of speaking / writing, the quantitative analysis does not show in fact an overuse of the gerund in comparison with other verb forms. A general positive prosody in the medium can also be documented through the use of both linguistic and paralinguistic devices.

Coming back to Swales, it can be finally concluded that Twitter is a genre used by an identifiable community with evident particular communicative purposes. Twitter has a number of specific characteristics affecting its schematic structure and style that result in regular and prototypical patterns which can account for differences in use and particularities of diverse groups of users due to special interests, culture or language.

ACKNOWLEDGEMENTS

We want to thank Ms. Jennifer Wilson (Linguistic Assessor for the British Council) for her revision and encouraging comments to the final draft of this paper.

REFERENCES

- Bazerman, C. (2000). Singular utterances: realizing local activities through typified forms in typified circumstances. In A. Trosborg (Ed.) *Analysing Professional Genres*. Amsterdam, John Benjamins.
- Boyd, D. M., and Ellison, N. B. (2007). “Social Network Sites: Definition, History, and Scholarship”. *Journal of Computer-Mediated Communication*, 13(1): 210-230. <http://dx.doi.org/10.1111/j.1083-6101.2007.00393.x>
- Cambridge University Press (1995). *Cambridge International Dictionary of English*. Cambridge, Cambridge University Press.
- Conrad, S., and Biber, D. (2000). Adverbial marking of stance in speech and writing. *Evaluation in text: Authorial stance and the construction of discourse*, 56–73.
- Java, A., Song, X., Finin, T. and Tseng, B. (2007) “Why we Twitter: Understanding microblogging usage and communities”. Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop.

- Kwak, H., Lee, C., Park, H. and Moon, S. (2010). "What is Twitter, a social network or a news media?" International World Wide Web Conference Committee (IW3C2) Raleigh, North Carolina, USA.
- Levy, P., S. Little and Aiyegbayo, O. (2007). Design for learning for the social *network generation: themes from a LAMS evaluation project*. In http://www.alt.ac.uk/altc2007/timetable/abstract.php?abstract_id=1274
- Misanchuk, M. and Anderson, T. (2001) "Building community in an online learning environment: communication, cooperation and collaboration". <http://frank.mtsu.edu/~itconf/proceed01/19.html>
- Mischaud, E. (2007). Twitter: Expressions of the whole self. An investigation into user appropriation of a web-based communications platform. London: Media@lse. Retrieved October 20, 2011 from http://www2.lse.ac.uk/media@lse/research/mediaWorkingPapers/MScDissertationSeries/Past/Mishaud_Final.pdf
- O'Reilly, T. (2004). "The architecture of participation". http://oreilly.com/pub/a/oreilly/tim/articles/architecture_of_participation.html
- O' Reilly, T. and Milstein, S. (2009). *The Twitter book*. Sebastopol, USA, O'Reilly Media Inc.
- Schmidt, H. (2009). *TreeTagger*. <http://www.ims.uni-stuttgart.de/~schmid/>
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford, Blackwell.
- Swales, J.M. (1990). *Genre analysis*. Cambridge, Cambridge University Press.
- Tseronis, A. (2009). *Qualifying standpoints. Stance adverbs as a presentational device for managing the burden of proof*. Utrecht, LOT Dissertation Series.
- Twitter. <http://twitter.com/> 8/11/2011
- Wikipedia <https://es.wikipedia.org/wiki/Emoticono> Retrieved October 1, 2011