

# The Emergence and Current State of Drug Trafficking-related Social Media in Mexico: A Text Mining Approach

*El surgimiento y situación actual de los medios sociales relacionados con el narcotráfico en México: Una aproximación desde la minería de textos*

Antonio Rico-Sulayes<sup>1</sup>

Manuela Gutiérrez-Leefmans<sup>2</sup>

Luis Villaseñor-Pineda<sup>3</sup>

## ABSTRACT

In 2010, a number of social media sites related to drug trafficking emerged in Mexico. A few of these sites gained popularity quickly and the number of sites proliferated. However, just a year later, some of the already popular websites closed after some of their users were killed. Based on figures resulting from text mining, this article describes the rapid emergence, success and dissolution of a number of drug trafficking-related social media sites. The study examines whether social media, especially mainstream platforms, have become an arena to discuss drug trafficking, or whether this type of conversation on social media is yet another victim of organized crime in this country.

**Keywords:** Computational Linguistics, Information Retrieval, Twitter, Organized Crime.

## RESUMEN

En 2010 una cantidad de redes sociales relacionadas con el tráfico de drogas surgieron en México. Algunos de estos sitios elevaron su popularidad rápidamente y el número de sitios proliferaron. De cualquier manera, solo un año más tarde algunos de los sitios más populares fueron clausurados después que algunos de sus usuarios fueron asesinados. Con base en las figuras resultantes del minado de textos, este artículo describe el veloz surgimiento, éxito y disolución de un número de redes sociales relacionadas con el tráfico de drogas. Los estudios examinan si los medios de comunicación, en particular las plataformas dominantes, se han vuelto una arena para discutir acerca del tráfico de drogas o que este tipo de conversación en los medios de comunicación sea otra víctima del crimen organizado en el país.

**Palabras clave:** lingüística computacional, recuperación de datos, Twitter, crimen organizado.

---

1 Departamento de Lenguas, Universidad de las Américas Puebla.

2 Departamento de Marketing, Universidad de las Américas Puebla.

3 Departamento de Ciencias Computacionales, Instituto Nacional de Astrofísica, Óptica y Electrónica.

## INTRODUCTION

The first specialized drug dealing-related social media sites appeared in Mexico almost a decade ago. Shortly after the emergence of these sites, some of them were thriving while others went unnoticed. After some of these sites gained great popularity, a few of their users were killed and consequently some of these websites closed. It is likely that because of this, all remaining sites eventually changed their conversation design. The options and tools that many of these sites offered for their users to initiate their own conversations were removed. However, as we will show in this paper, the conversation on social media about drug dealing in Mexico, especially self-regulated conversation, is still taking place. It has only migrated to other venues.

In order to reach the conclusion that self-regulated conversation about drug trafficking in Mexico, although much reduced, is still present on social media, we will present an analysis of data from both emergent specialized drug dealing-related social media sites as well as from the current conversational stream of the mainstream platform Twitter. By presenting the first type of data, we will argue that, although some initial sites did disappear due to lack of interest from their users, some went down in the midst of their increasing popularity. Therefore, whether their administrators explicitly acknowledge it or not, it can be assumed that these sites were also victims of organized crime. After analyzing current data from mainstream social media, we will show that where expressions that have been characterized as *narco language* (Saldívar, 2014) are present, the conversation is about drug dealing. This is true in virtually all tweets that contain this type of language and they are placed in the context of Mexico. Despite the fact that many of the words included in the *narco lexicon* we scraped from Twitter also have other more general meanings in everyday language, when geographically referring to Mexico, between 96.4% and 97.7% of the messages that contain them talk about drug dealing and organized crime activities. This shows that the conversation about drug dealing on social media in this country is not only vibrant but enjoys ongoing popularity.

## DATA COLLECTION AND METHODOLOGY

In order to lay out the development of drug dealing-related social media in Mexico, this study will present data from two different chronological periods. First, we will show data from some of the most successful sites that appeared during the emergence of this internet phenomenon in 2010. When talking about this period, it is important to note that the sites discussed were solely devoted to the topic of drug trafficking and offered tools for self-regulated conversations among their users. The data presented will include figures about the number of conversations, individual contributions, active users, and number of lexical tokens. These numbers will be reported for several points in time during a two to five-year period since the birth of these social media sites to introduce their early development. Since many of these sites have disappeared and all of them have changed their conversation format, we will show current figures that result from scraping Twitter. In the scraping of Twitter posts, we have harvested tweets containing individual words, multi-word lexical units, and hashtags that belong to two semantic fields: the names of drug cartels and the names given to murder victims. These semantic fields have been inspired by research in *narco language*, an expression of *narco culture* that has been studied using specialized corpora in Saldívar-Arreola (2014). For our study, human beings annotated the

retrieved tweets in order to judge their relevance. Tweets were annotated if they met two criteria: they referred geographically to Mexico and they were related to the topic of drug dealing or organized crime activities. By doing this, we demonstrate that mainstream social media platforms have become an important venue for communication about drug dealing in this country. It should also be noted that we chose Twitter to explore the current status of mainstream social media in the conversation about drug dealing because, although the original sites specifically devoted to this topic either disappeared or changed their conversational tools, Twitter has made its data available to the research community to study trends in conversation. This has resulted in a plethora of studies in the computational linguistics and text mining communities that are based on this type of data. The present article aims to add to this body of research.

#### USAGE ANALYSIS OF SPECIALIZED DRUG DEALING-RELATED SOCIAL MEDIA

Specialized drug dealing-related social media websites emerged and became popular in Mexico in 2010. The first author of this article has been monitoring these sites for almost a decade (Rico, 2011; 2018). A Google query in September 2010, with the keywords *blog*, *narco* (drug dealer / drug dealing), *narcotráfico* (drug dealing) and *México*, would return a good number of these sites among its hits. Visiting these sites back then, the majority of them had only a few dozen usernames registered and most of these usernames had not posted any comments. But this was not the case for all sites. The proliferation of these sites was likely prompted by the success of arguably the most popular drug dealing-related blog at the time: *El Blog del Narco*. This site was created on March 2, 2010 and promptly captured the attention of international media (for example CNN, 2010). Just a few months after its creation, *El Blog del Narco* included contributions by alleged average citizens, relatives of the kidnapped and murdered, policemen, members of the military, and drug dealers. A number of months after their creation, many of the other drug trafficking-related sites, which most likely emerged as a result of the success *El Blog del Narco*, vanished or were abandoned by their creators and casual users (Rico, 2012).

However, not all drug dealing-related social media sites that disappeared during the emergence of this phenomenon did so because of a lack of interest from their audiences. This was the case of the online forum *Frontera al Rojo Vivo* (2011), whose name means *White-Hot Border*. This online forum was created and maintained by *El Norte* (The North), a local newspaper belonging to a nationwide print media company (elnorte.com, 2018). *El Norte* decided to close their online forum after a number of *abuses* and *personal attacks* involving forum users took place (*Frontera al Rojo Vivo*, 2011). These comments, in what remained of the forum shortly after its deactivation, were vague but likely alluded to the murder of two social media users in the city of Nuevo Laredo. One of the victims was a man whose body was hung from an overpass. He was killed “for snitching in *Frontera al Rojo Vivo*”, as could be read in a note attached to his leg (Goodman, 2011, p. 5). A banner hung from the overpass next to the man’s body also threatened “all the Internet snitches” in *El Blog del Narco*, formerly mentioned, and *Denuncia Ciudadana*, a website created at the time by the *Política Federal Mexicana* for citizens to give anonymous tips (Procuraduría General de la República, 2010). The other victim that might have prompted *El Norte* to close its forum was a woman. Her body was also hung from a pedestrian overpass. She was tortured and killed because of her posts in *NuevoLaredoEnVivo*,

another popular social media website that was created during the emergence of this phenomenon. This was made explicit by a sign left by the murderers next to her body (*Borderland Beat*, 2011; Goodman, 2011; Stevenson, 2011). These murders were only one manifestation of a series of continuous threats against online media users and anonymous tipsters by drug cartels (*Borderland Beat*, 2011). These early acts made the risks and challenges of maintaining drug dealing-related social media sites in Mexico explicit.

As it relates to the usage figures of drug dealing-related social media sites during their emergence, six months after the creation of *El Blog del Narco*, the first author of this article started monitoring, harvesting, and processing the posts of its associated online forum. The focus on this type of social media site, online forums, was influenced by the perspective of Owyang (2008). According to this author, online forums are a peculiar form of conversational text. For Owyang, online forums are a form of social media that has as its main goal constructing a multi-party conversation. The peculiarity of online forums is that they offer an opportunity for communication among many users of equal status, that is, their topic threads are intended to promote a conversation among many users with little guidance. In contrast, when dealing with blogs, the *speaker* or blogger leads the interaction between him or herself and audience members (Miller and Shepherd, 2004; Owyang, 2008). Online forums are, however, *social mixers* in Owyang's words, and therefore allow their users to have self-regulated conversations.

In the course of monitoring the online forum associated with *El Blog del Narco*, several pieces of data were registered at different points in time. These data included the raw number of messages posted, as reported by the site, and the number of topic threads in which these messages appeared. After harvesting the entire forum with a web crawler several times, all messages were preprocessed to count the number of messages posted by signed or logged-in users, along with the total number of these users. Table 1 below shows figures at three points in time: the first two, near the beginning of the interaction with the forum, and the last one, just a few weeks before the online forum was deactivated by its administrators. The figures in Table 1 include, from left to right, the raw number of messages posted (as reported by the site), the number of topic threads, the number of messages posted by signed users, the number of signed users who posted comments, and the total number of words or tokens in all messages by signed users.

**Table 1.** Usage figures for a popular drug dealing-related online forum.

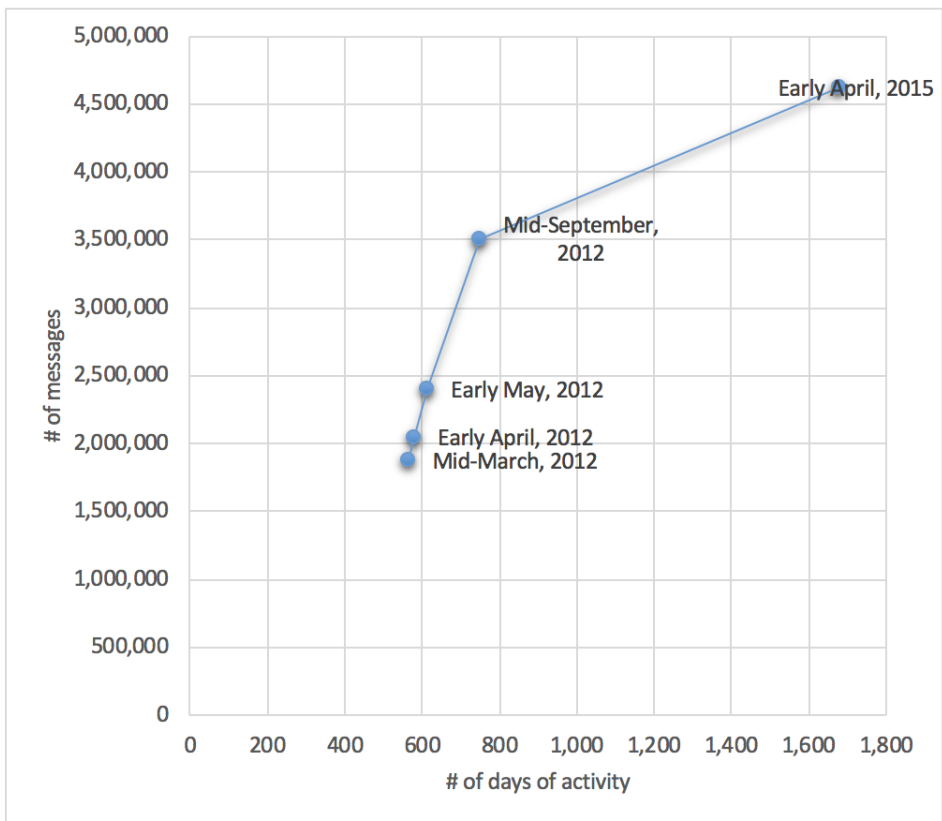
Time period	Raw # of posts	# of topic threads	# posts by signed users	# of active users	# tokens by signed users
Early September, 2010	41,571	4,205	37,571	1,026	2,128,049
Late October, 2010	61,304	6,624	-	-	-
Early July, 2011	130,257	-	119,345	2,579	8,872,841

Although not all figures are included for every reported period —the numbers were registered only as they were needed for other research purposes— the data in Table 1 allows us to see the growing trend in online forum contributions. As shown in the second column, from left to right, the raw number of messages more than tripled in just 10 months, from September 2010 to July 2011, as well

as the number of messages posted by signed users (fourth column). This last piece of information is even more significant because the first figure may include duplicated messages, as well as messages posted by anonymous, casual users or *bots*; however, the second figure shows preprocessed data figures, where these types of messages were eliminated. Using only the remaining messages, the number of words or tokens produced by signed users was calculated and included at the rightmost column. This figure increased over four times in the examined period. This reveals that not only did the number of messages increase, but the messages themselves became longer. These data demonstrate the popularity this type of social media had, which also leads to the assumption that the eventual deactivation of this online forum, a few weeks after the last crawling in mid-July, was not due to the lack of interest of its users.

Another important social media website monitored during the emergence of this internet phenomenon was the blog named *NuevoLaredoEnVivo* (Nuevo Laredo Live), which was formerly mentioned in relation to the two users killed by members of organized crime. This site was created in 2010 and, similar to *El Blog del Narco*, gained popularity very quickly. Although it was originally conceived as a blog, *NuevoLaredoEnVivo* also offered a chat room starting in September 2010. The tool allowed users to have self-regulated conversations and therefore exercise greater control over topics. This form of communication was even more ephemeral than conversations in online forums. The number of contributions in this site's chat room increased rapidly, as can be seen in Figure 1 below. This figure plots the number of contributions, on the *y* axis, against the number of days of activity, *x* axis, for this chat room. This last piece of information was also displayed on the website.

**Figure 1.** Usage data of a popular chat room.



Although the registration of usage figures for the chat room of *NuevoLaredoEnVivo* began eighteen months after its opening, the trend line in Figure 1 illustrates that its chat room had a large increase in popularity in 2012. Over the course of the six months covered by the first four observations, from March to September, the number of messages posted in the chat room almost doubled. This results in a steep trend line for this time period. This high activity must have slowed down at some point, as three years later there were only over 1.1 million new contributions compared to over 1.6 million added solely during the six-month period previously mentioned. This decrease in activity is represented by the reduction in the steepness of the trend line for this three-year period. It is hard to say why the popularity of this site decreased, however, it is clear that there have been important changes in the design of the site. Beyond the several URLs, or internet addresses, to which *NuevoLaredoEnVivo* has migrated throughout the years (2012; 2015; 2018), the structure of user conversation for this site is different now. Currently the site offers three different chat rooms to make reports to different authorities. One has the purpose of reporting incidents to the city mayor's office in Nuevo Laredo, a second serves the purpose of requesting police assistance and the last one is used to report dangerous situations. None of these chat rooms show the number of contributions, and any off-topic posts are deleted by administrators. Clearly, the conversation among users in *NuevoLaredoEnVivo* is not self-regulated anymore; neither do the site tools promote multi-party exchanges.

We could speculate that both, the disappearance of the drug dealing-related social media sites, as well as the dramatic changes in the conversational tools and capabilities of the remaining sites, negatively affected the conversation about drug trafficking and organized crime through social media in general. To explore if this is the case, we decided to further investigate other mainstream social media platforms, in this case Twitter. We wanted to see if mainstream platforms like this had anything to offer to their social media users or if they were considered too rigid to promote self-guided conversation about these kinds of sensitive topics. The next section presents our research on mainstream platforms regarding users' conversations on drug dealing.

#### TEXT MINING ANALYSIS OF TWEETS ABOUT DRUG DEALING IN MEXICO

The use of social media data for scientific research has become ubiquitous (Danneman & Heimann, 2014), not only in the field of computational linguistics, but in the much wider data science community. The specific use of tweets in text mining covers a great range of applications, from sentiment analysis on human crises (Öztürk & Ayvaz, 2018), to the profiling of social media users regarding age and gender (Álvarez et al., 2018). Although the number of specific applications is too large to list here, this article adds to this wide range of applications the topic of drug dealing in Mexico.

After scraping tweets about our topic for one week, we examined them following a simple relevance analysis. In the general field of information retrieval, the list of hits returned by a search engine after running a query can be analyzed with two complementary measures that address the relevance of results, precision and recall (Manning et al., 2009). Precision is the proportion of hits or documents that are relevant for the information need expressed through the query. For example, one could run the query *drug dealing social media* looking for websites that allow users to interact with each other sharing information

on drug dealing. In this case, the precision of the search engine, say Google, is equal to the number of webpages that match the former description among all the websites listed in the hits divided by the total number of hits. If Google reports 200 hits and looking at all of them, only 40 are the type of website one is looking for, the precision in this case is 20% (i.e.,  $40/200 = 0.2$ ). In the context of this paper,

$$Precision (P) = \frac{\# \text{ relevant tweets}}{\# \text{ all tweets retrieved}}$$

so we will calculate this measure using this formula. A related measure, recall, is equal to the proportion of all relevant documents retrieved with the query. The calculation of recall requires knowledge of all relevant documents in the searched space. This implies knowing all relevant pages in the entire internet, in the case of Google, or all relevant tweets in Twitter's stream of data. Since this condition is very difficult to achieve, recall is usually applied in more restricted collections of data where the entire contents have been previously annotated for some other purpose. In this study, we will look only at the precision of the database obtained by scraping Twitter. We will not do this to judge the efficiency of the computational tools used to scrape Twitter's stream of data, but to argue whether or not the topic of drug dealing in Mexico is part of the conversational trends on Twitter.

In order to scrape Twitter, we began by looking at the study on narco language in Saldívar (2014). Trying to identify the lexical elements that distinguish the language of drug dealing, this author assembles a specialized corpus with documents from six genres related to this topic. The six genres are 1) online blogs, 2) literature, 3) essays, 4) music lyrics, 5) press materials and 6) interviews with people whose job is related to this phenomenon, from police officers to drug dealers. By making a statistical comparison of this text collection against a reference corpus, the author produces a list of 543 lemmatized forms that characterize the lexicon of drug dealing. He then divides these words in a number of semantic fields in his lexicological analysis. Saldívar-Arreola (2014) also gives a list of the ten most common words in his specialized lexicon of narco language: *sicario* (hitman), *mariguana* (marijuana), *narco* (drug dealer), *coca* (cocaine), *narco* (drug dealing), *Zetas* (the proper name of a cartel), *malandrín* (scoundrel), *compa* (buddy / comrade), *ejecutado* (person shot execution-style), and *cuerno de chivo* (detachable-magazine assault rifle).<sup>4</sup> Several of these words can be grouped in semantic fields, such as names of gang members (*sicario* hitman, *narco* drug dealer, *malandrín* scoundrel, and *compa* buddy / comrade) and drugs (*mariguana* marijuana and *coca* cocaine). However, we were particularly interested in two words (*Zetas* and *ejecutado*) that point to two other semantic fields: proper names of cartels and terms to refer to the murder victims. We were interested in these two semantic fields because we wanted to explore whether users tweeted about topics that may be considered dangerous for the tweeter (in the former section we mentioned the deadly attacks against social media users) or, at least, sensitive to talk about. In order to explore the presence

<sup>4</sup> *Cuerno de chivo* literally means *goat horn* because of a metaphor between the horn of this animal and the curved shape of the magazine in many assault rifles, such as the AK-47.

of conversation having to do with these two semantic fields, we scraped Twitter for one week during which we harvested tweets that contained the proper name of one of five cartels or that included one of five terms that refer to victims murdered using different methods. These were selected because they stood out, after having run a number of queries in the search field of Twitter, of the most frequently occurring keywords.

The proper names of cartels were harvested using both a hashtag and the words in the name itself. When using a hashtag, the octothorp, (#), commonly known as the pound sign, was followed by either the name itself in camel case (all words collapsed into one with their first letter capitalized) or its acronym, e.g, #NameOfCartel or #NOC. Table 2 below shows the results of scraping the five cartel names. The number of relevant tweets in column four from right to left met our two criteria of relevance; they referred geographically to Mexico and talked about drug dealing. For example, in the line immediately below the header, the scraping of the hashtag composed by the octothorp followed by the acronym of the first cartel name retrieved 494 tweets in over a week. After a human annotator analyzed all these tweets, she decided that all of them were relevant, meeting the two criteria. Thus, the precision of the scraping for this hashtag was 100%.

**Table 2.** Twitter scraping precision for names of cartels.

Target content	Retrieved	Relevant	Precision
Hashtag with cartel 1 acronym	494	494	1.000
Cartel 1 name	47	47	1.000
Hashtag with cartel 2 name in camel case	0	-	-
Cartel 2 name	75	33	0.440
Hashtag with cartel 3 name in camel case	0	-	-
Cartel 3 name	44	44	1.000
Hashtag with cartel 4 name in camel case	27	27	1.000
Cartel 4 name	1	1	1.000
Hashtag and cartel 5 name in camel case	5	5	1.000
Cartel 5 name	2421	2391	0.988
	Average of averages		0.928
Total number of tweets	3114	3042	0.977

As can be seen in Table 2, out of all ten scraping scripts only two did not retrieve any tweets, those being the ones with hashtags for the cartel 2 and cartel 3 names. The precision for these two cases was not calculated (mathematically that is not feasible) and they were excluded from the calculation of the precision average. As for this figure, the average of averages in the second-to-last line was high, 92.8%, but it was negatively influenced by the cartel 2 name, which is composed of words that are commonly found in Spanish. As the scraping retrieved the tweets that contain the searched words in any order, scraping the words in this cartel name produced a low number of hits in this case. However, the general average of all tweets reduced this isolated negative effect giving an even higher overall precision of 97.7%.



As to the other semantic field explored, the one consisting of terms to refer to murder victims, the words in this case were scraped in two morphological forms, singular and plural. There was no need to look for feminine forms in Spanish because all these terms are psychologically associated with the word *cadáver* (corpse/ dead body), which is masculine in this language. Along with their plurals, the five terms that were scraped from Twitter are *ejecutado* (person shot execution-style), *embolsado* (bagged dead body), *encobijado* 'dead body wrapped in blanket', *encajuelado* (dead body left in trunk of abandoned car), and *pozoleado* (dead body in barrel with caustic liquid).<sup>5</sup> Table 3 below shows the results of scraping these terms for one week on Twitter.

The first difference to note between Table 3 and 2 is that the third column from the left in Table 3 shows the number of tweets that geographically refer to Mexico. Since Table 2 above included the results of looking for names of Mexican cartels, very frequently the tweets containing them also referred to this country. However, in the case of the words reported in Table 3, several of them are commonly used in other countries. Due to this, to calculate precision we used the modified formula  $P = \# \text{ relevant tweets} / \# \text{ all tweets retrieved relevant for Mexico}$ . This is because we wanted to see how often tweeters were talking about drug dealing and organized crime activities as compared to all the times they used these words to refer to Mexican context.

**Table 3.** Twitter scraping precision for terms referring to murder victims.

Target content	Retrieved	Rel Mexico	Relevant	Precision
Encajuelado	3	3	2	0.667
Encajuelados	0	-	-	-
Encobijado	55	51	45	0.882
Encobijados	18	16	14	0.875
Pozoleado	0	-	-	-
Pozoleados	0	-	-	-
Ejecutado	1949	317	290	0.915
Ejecutados	9649	1754	1712	0.976
Embolsado	146	68	66	0.971
Embolsados	31	16	16	1.000
	Average of averages			0.898
Total number of tweets	11674	2225	2145	0.964

As seen in Table 3, of all five terms scraped, only one did not retrieve any tweets with either its singular or plural form: the word *pozoleado* (dead body in barrel with caustic liquid). As we analyzed the data, we found a possible explanation being that this term is currently competing with other forms such as *empozolado*, which is just a morphological variation, and *entambado* that literally means 'put in a barrel.' As before, all lines with zero results were excluded from the precision average. As for the average of averages, 89.8%, it was negatively influenced by *encajuelado* (dead body left in trunk of abandoned car), which

<sup>5</sup> The word *pozoleado* is derived from *pozole*, a large kernel corn, and by extension, the soup made from it. Pozole soup is usually made with chopped meat and vegetables. It may be the thick, chunky consistency of this soup that prompted its comparison with a barrel containing a dissolving dead body.

obtained a 66.7% precision. For the complete semantic field, however, the general average of all tweets also reduced the negative effect and rendered an overall precision of 96.4%, which is slightly lower than the precision of the semantic field of proper names of cartels. It should be noted that the results for the terms in Table 3 are especially significant because several of these terms have other, more general meanings in the language. Thus, *ejecutado* can mean (executed / performed), *embolsado* (bagged), *encobijado* (wrapped in blanket), and *encajuelado* (kept in a car's trunk). The only neologism in the list is actually *pozoleado*.

#### RESULTS AND INTERPRETATION

As mentioned at the beginning of this article, besides the analysis of the emergence of specialized drug dealing-related social media, our general objective has been to explore the current status of mainstream social media in the conversation about drug dealing. Specifically, since all specialized social media sites that appeared almost a decade ago have either disappeared or changed their tools, self-regulated conversation is no longer promoted on their sites. We wanted to see whether Twitter has become a venue for talking about drug dealing in Mexico. After scraping five names of frequently mentioned cartels in Twitter for a week, we were able to harvest 3,114 tweets. As mentioned before, 97.7% of these tweets both talked about drug dealing or organized crime activities and referred to Mexico. This occurred despite the fact that half of the scraping tools looked for tweets that contained words in free order. The scraping of five terms to refer to murder victims retrieved more than 11,674 hits. Out of these, 2,225 tweets geographically referred to a Mexican context. Looking at these Mexico-related tweets, 2,145 talk about drug dealing. This means that, despite the fact that all but one of the terms have other more general meanings in the language (beyond their meaning as spine-chilling corpses), 96.4% of the times tweeters used these words to refer to a Mexican context, their topic was also drug trafficking or other organized crime activities. This clearly demonstrates that mainstream social media has become a venue for its users to express their thoughts about this topic. This has happened regardless of the fact that self-regulated conversation about drug trafficking in Mexico was originally negatively affected by various factors that forced specialized sites to either close or change their conversational design.

#### CONCLUSION

Before finishing, we want to mention that there are currently several specialized social media sites devoted to the topic of drug dealing in Mexico and their popularity appears to be high. In this article, we have tried to emphasize that the majority of such sites have changed the format of their tools, such as excluding self-regulated conversation from their design. An example of a still very popular site is *El Blog del Narco* (2018). Using a web monitoring service (Alexa, 2018), we found, in June 2018, that this site had 2,579,850 monthly page views from 57,187 unique visitors. With its gruesome photos and videos (which include decapitations and other atrocities), this site obtains 48% of its visitors from the United States and 39.9% from Mexico. Despite its popularity, *El Blog del Narco's* online forum was removed in 2011. Another example of a current popular site is *NuevoLaredoEnVivo* (2018), which has not been hosted on its current server for very long (we estimate that it has been on its current server for less than a year), and yet, it already reports over 4.5 million page views in

early August, 2018. The continued popularity of sites of this kind, despite their changes in the tools they offer users, testifies to the continued need for the type of data and analysis presented in this paper.

#### ACKNOWLEDGEMENTS

We want to thank senior students Jessica Lizbeth Sánchez Villa and Viviana Catalina Ríos Ortega for the many work-study hours they invested in various tasks including human annotation.

#### REFERENCES

- Alexa. (2018). Retrieved from <https://www.alexa.com>
- Alvarez, M. A., Pellegrin, L., Montes-y-Gómez, M., Sánchez-Vega, F., Jair Escalante, H., López-Monroy, A. P., Villaseñor-Pineda, L., Villatoro-Tello, E. (2018). A visual approach for age and gender identification on Twitter. *Journal of Intelligent and Fuzzy Systems*, Vol. (34), pp. 3133-3145.
- CNN. (2010) *Television broadcast*. Retrieved from <http://edition.cnn.com/video/#/video/world/2010/08/19/romo.mexico.narco.blog.cnn>
- Borderland Beat. (2011). Nuevo Laredo, the Silent War. Retrieved from <http://www.borderlandbeat.com/2011/09/nuevo-laredo-silent-war.html>
- El Blog del Narco*. (2018). Retrieved from <https://elblogdelnarco.com/>
- elnorte.com. (2018). Retrieved from <http://www.elnorte.com/>
- Frontera al Rojo Vivo*. (2011). Retrieved from <http://gruporeforma.elnorte.com/libre/offlines/mty/frontera/>
- Goodman, J. D. (2011). In Mexico, Social Media Become a Battleground in the Drug War. *The New York Times: The Lede*. Retrieved from <http://thelede.blogs.nytimes.com/2011/09/15/in-mexico-social-media-becomes-a-battleground-in-the-drug-war/>
- Danneman, N., & Heimann, R. (2014). *Social Media Mining with R: Deploy Cutting-Edge Sentiment Analysis Techniques to Real-World Social Media Data Using R*. Birmingham, UK: Packt.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge, England: Cambridge University.
- Miller, C. R., and Shepherd, D. (2004). Blogging as Social Action: A Genre Analysis of the Weblog. In L.J. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, and J. Reyman (Eds). *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*. Retrieved from [http://blog.lib.umn.edu/blogosphere/women\\_and\\_children.html](http://blog.lib.umn.edu/blogosphere/women_and_children.html)
- NuevoLaredoEnVivo*. (2018). Retrieved from <https://nuevolaredoenvivo.webnode.es/>
- NuevoLaredoEnVivo*. (2015). Retrieved from <http://nuevolaredoenvivo.blogspot.mx/>
- NuevoLaredoEnVivo*. (2012). Retrieved from <http://nuevolaredoenvivo.es.tl/>
- Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, Vol. (35), pp.136-147
- Owyang, J. (2008). Understanding the Difference between Forums, Blogs, and Social Networks. *Web Strategy*. Retrieved from <http://www.web-strategist.com/blog/2008/01/28/understanding-the-difference-between-forums-blogs-and-social-networks/>
- Procuraduría General de la República. (2010). *Denuncia Ciudadana*. Retrieved from <http://www.pgr.gob.mx/denuncia/denuncia.asp#>
- Rico, A. (2018). *Authorship Attribution on Crime-Related Social Media: Research on the darknet in forensic linguistics*. Rome: Aracne.
- Rico, A. (2012). Quantitative Authorship Attribution of Users of Mexican Drug Dealing Related Online Forums. Unpublished doctoral dissertation, Washington, DC., George-

- town University. Retrieved from [https://repository.library.georgetown.edu/bitstream/handle/10822/557726/RicoSulayes\\_georgetown\\_0076D\\_11981.pdf](https://repository.library.georgetown.edu/bitstream/handle/10822/557726/RicoSulayes_georgetown_0076D_11981.pdf)
- Rico, A. (2011). Statistical Authorship Attribution of Mexican Drug Trafficking Online Forum Posts. *International Journal of Speech, Language and the Law*, vol. (18), pp. 53-74.
- Saldívar, A. (2014). *Análisis Lexicológico del Narcolenguaje en Baja California*. México, DF: Universidad Autónoma de Baja California.
- Stevenson, M. (2011). Woman Decapitated in Mexico for Web Postings. *NBCnews.com*. Retrieved from [http://www.nbcnews.com/id/44657254/ns/world\\_news-americas/t/woman-decapitated-mexico-web-postings](http://www.nbcnews.com/id/44657254/ns/world_news-americas/t/woman-decapitated-mexico-web-postings)