

SaveYourInternet: Exploring polarities and actions in the discourse of Twitter's posts about Article 13's approval

SaveYourInternet: Explorando polaridades y acciones en el discurso de publicaciones en Twitter relacionadas con la aprobación del Artículo 13

José Vicente Benavent Cháfer

Universitat de Barcelona

jbenchbe7@alumnes.ub.edu

Resumen: Al contemplar el auge en uso y popularidad de las redes sociales para comunicarse y, en definitiva, compartir opiniones, no es de extrañar que investigadores de diferentes especialidades se interesen en analizar dichas interacciones online. Más allá de su aplicación comercial, el presente estudio pretende explorar el potencial de las técnicas de análisis de sentimiento junto con la lingüística de corpus y la gramática sistémica funcional. Concretamente, se usa la teoría de la transitividad de Halliday y Matthiessen para discernir cómo se aborda la aprobación del Artículo 13 por la comunidad de Twitter desde un punto de vista lingüístico, principalmente enfocado en los verbos y agentes involucrados en el discurso. Dicha reforma legal pretende regular el uso de materiales protegidos por derechos de autor en plataformas de internet como YouTube y Facebook, entre otras. De este modo, se investiga si la diferencia entre textos producidos durante días diferentes abordando el mismo tema presenta diferencias significativas hasta el punto de considerarse géneros discursivos distintos.

Palabras clave: análisis de sentimiento, teoría de la transitividad, lingüística de corpus

Abstract: As the use of social media grows in users and popularity to communicate and share opinions, researchers from various fields are, unsurprisingly, interested in analysing such online interactions. Beyond its commercial application, sentiment analysis is consulted in the present research to explore the potential its combination with corpus linguistics and systemic functional grammar. Specifically, Halliday and Matthiessen's transitivity theory is employed to recognise how the Twitter community tackles Article 13's approval from a linguistic point of view, mainly focused on the verbs and agents involved in the discourse. The aforementioned legal reform attempts to regulate the use of copyrighted material in online platforms like YouTube and Facebook, amongst others. Thus, this paper investigates whether significant linguistic differences arise across the resulting texts addressing the same issue in the span of a few days, up to the point that they become distinct discursive genres.

Keywords: sentiment analysis, transitivity theory, corpus linguistics

1. Introduction

The widespread use of microblogging sites like Twitter has motivated research on converging fields related to communication technologies and online information sub-fields. Given the large quantities of data available in such platforms, researchers tap into the so-called opinion mining techniques to survey their target population. More specifically, sentiment analysis is currently undertaken at a large scale by local and global companies to gauge the likeability of their products and services by consulting the data generated by potential customers (social media users), thus adapting commercial marketing strategies as a result (Agarwal *et al.*, 2011: 30). Due to its time-consuming nature, manual sentiment analysis is overtly avoided in the above-mentioned scenarios. Consequently, automatic alternatives assume most of the investigation workload (Chikersal *et al.*, 2015: 50).

As for the present paper, sentiment analysis tools (such as Chorus Analytics) are employed as a data-gathering method. In other words, performing an inherently machine-based sentiment analysis is disregarded in the current study, even if the analytical section relies on the founding principles of the aforementioned discipline. Instead, this research advocates for the manual examination of lexis through the concordancing software AntConc, as opposed to the traditional automated algorithm-driven option. Once the linguistic content is retrieved, the analysis shall proceed to classify it according to transitivity theory, a taxonomy centered around types of verbs and their defining features (Halliday and Matthiessen, 2014).

In essence, this study sets out to explore the discursive strategies employed by Twitter users when discussing possible legal and ethical implications stemming from Article 13's hypothetical approval. This reform deals with copyright infringement concerning audiovisual material being uploaded to content sharing service providers such as YouTube and Facebook. Not only is SaveYourInternet a social movement, it is also a communicative exchange worthy of examination, for scrutinising the public's opinions, notions of agency, and verbal forms may create a more comprehensive picture of the layperson's view on policymaking, as far as online communications are concerned. As Taboada (2016: 27) points out, interpreting the polarity of a clause does not only entail examining isolated lexical units, but the whole discursive structure and argumentation may end up revoking the polarity initially expressed. By analyzing a set of tweets retrieved in a span of 12 days around Article 13's alleged approval (17-28 March 2019), it is sought to find out whether outstanding differences in polarities across different texts and time periods may end up revealing distinct discursive genres altogether, or rather distinct stages thereof.

2. Theoretical frameworks

2.1. Sentiment analysis

In computational linguistics and NLP (Natural Language Processing), the notion of 'sentiment' may vary across sub-disciplines. However, the area of interest in this research is that concerned with 'opinion mining', a classifying concept that divides evaluative/affective components of speech into two distinct polarities: either negative or positive (Pang and Lee, 2008: 6). Some researchers investigate the potential of this technology applied in large online domains for eliciting the public opinion with regards to policy-making and ultimately guide the rulemaking process (Cardie *et al.*, 2006; Kwon *et al.*, 2006). Even with the inclusion of more refined techniques like *kernel* or *unigram models* (Agarwal *et al.*, 2011), such projects hinge on machine-learning analyses in their entirety, thus neglecting the importance of human interpretation of the studied lexis. Indeed, researchers do not deem appropriate a semantic analysis that overlooks linguistic properties such as word order (Socher *et al.*, 2013: 1633). In addition to this, the aforementioned consideration could be extended at the discourse level due to the context-dependent nature of interpretative works on semantics. In this paper, sentiment analysis is undertaken to offer an overall perspective on semantic orientation, instead of conducting the linguistic/affective analysis per se.

2.2. Transitivity theory

Far from its traditional syntactical conception, Halliday and Matthiessen's (2014: 220) notion of 'transitivity' is centered around the meaning-making of a clause which is embedded into three semantic metafunctions occurring simultaneously, namely the 'textual' (clause as a message), the 'interpersonal' (clause as a social exchange), and the 'ideational' (clause as a representation of experience) function. This framework accounts for the models of thought or schemata which construct the reality around us: the agents, actions/processes, and the co-occurring circumstances (2014: 213). Such elements vary in their wording according to the major types of processes: the 'material', 'behavioural', 'mental', 'verbal', 'relational' and 'existential' types (Halliday, 1970: 213).

Process type	Category Meaning	Participants
<u>Material:</u> Action Event	'doing' 'doing' 'happening'	Actor- Goal
<u>Behavioural</u>	'behaving'	Behaver
<u>Mental:</u> Perception Affection Cognition	'sensing' 'seeing' 'feeling' 'thinking'	Senser- Phenomenon

<u>Verbal</u>	'saying'	Sayer – Verbiage-Target/Receiver
<u>Relational:</u> Attribution Identification Possession	'being' 'attributing' 'identifying' 'possessing'	Token- Value Carrier- Attribute Identified- Identifier Possessor- Possessed
<u>Existential</u>	'existing' 'happening'	Existent

Figure 1: Summary of process types and sub-types according to Halliday's transitivity theory (1985: 131).

As shown in Figure 1, material processes are related to verbs of 'doing' or 'happening'. Action types are normally transitive verbs (targeting a direct object), whereas event types tend to be intransitive. As for the participants involved, they are typically represented by nominal clauses, as shown in the examples provided in Figure 2 below.

a)	The hunter	ran	b)	The hunter	caught	the prey
	Actor	Process		Actor	Process	Goal (Direct object)
	nominal clause	verbal clause		nominal clause	verbal clause	nominal clause

Figure 2: Material processes sub-types: Events (a), and Actions (b).

Behavioural processes, on the other hand, generally refer to verbs which a) are intransitive (hence having one participant/Behaver) and b) describe an action in which both physical and mental components cannot be separated. For instance: His friend is always *laughing* (Halliday and Matthiessen, 2014: 301).

The following row contains mental types, which allude to the realm of senses, or 'sensing'. As hinted in Figure 1, there are four distinct sub-types: perception (seeing, hearing), affection (liking, feeling), cognition (thinking, pondering), and desiderative mental processes (want, wish). Instead of an Actor and a Goal, mental processes add Sensers and Phenomena as participants, as Figure 3 illustrates below.

a)	My sister	likes	the gift	b)	The gift	Pleased	my sister
	Senser	Process	Phenomenon		Phenomenon	Process	Senser
	nominal clause	verbal clause	nominal clause		nominal clause	verbal clause	nominal clause

Figure 3: Mental process: affection (differing word order).

Since Sensers/Phenomena are not constrained by syntactic order, the former are defined as conscious living entities whereas the latter may refer to objects, actions, or facts.

Verbal processes are concerned with 'saying' and 'meaning', as participants are Sayers (entities who speak), and Receivers/Targets (subjects that receive the action), occasionally including Verbiage (a term alluding to the message or words themselves) (Halliday and Matthiessen, 2014: 302).

Relational types refer to states of 'being' through the copulative verb *to be* or linking verbs (seem, appear) (Halliday and Matthiessen, 2014: 286). The general pattern construction for such clauses are Tokens ascribed with Values, thus breaking down relational types into three distinct sub-types: attribution, identification, and possession. As the name suggests, attribution refers to matching a carrier, or subject of a clause (Mark) with an attribute (smart). Identification sub-types are also coded as "circumstantial" (Halliday and Matthiessen, 2014: 292), since the Identified discloses information as of *who* (He), whilst the Identifier relates to *where/with whom/in which manner* (is at the park). Finally, the possession sub-process establishes a relationship of ownership between two components within the clause: The Possessor (He) and the Possessed (owns a house), often employing the Saxon's genitive (Tanya's) (Halliday and Matthiessen, 2014: 298).

Lastly, existential processes are exemplified with verbal forms equivalent to 'existing' or 'happening'. Verbal clauses in this domain may either use the verb *to be* or a standalone form (occur). A nominal clause usually fulfills the role of Existent (an incident), which is later materialised through existential verbs (occurred, happened) (Halliday and Matthiessen, 2014: 308).

2.3. Legal framework

To briefly explain the intricacies of Article 13 and its legal implications, this paper shall refer to the latest published work by the European Parliament, the briefing on copyright in the digital single market, published on 22 March 2019¹. The aforementioned article is comprised in a "legislative package" intended to adapt the existing EU legal framework on copyright regulations and its fair use to the conditions of the digital age (Madiega, 2019: 1). The premise of this reform lies in safeguarding the authors' rights to receive an appropriate remuneration for the exploitation of their work in the internet through licensing agreements, although some cases are not subject to these specifications, such as their use in educational settings for learning purposes, data mining operations (scientific research), preserving the cultural heritage, and material being used as caricatures, parodies, pastiches or quotations (Madiega, 2019: 11). Even though the implementation of a more sophisticated filter for detecting unauthorised works in uploaded content was extensively discussed (such as YouTube's Content ID), the final text states that no

¹ As this document notes, this briefing "should not be taken to represent an official position of the Parliament", since it is intended to "assist [the members and staff of the European Parliament] in their parliamentary work" (Madiega, 2019: 14).

monitoring is required, in order to comply with the E-Commerce Directive and the European Charter of Fundamental Rights. Instead, content sharing service providers shall implement “an effective and expeditious complaint and redress mechanism that is available to users” (Madiega, 2019: 10).

3. Methodology

The rationale of this study for using a synergy of systemic functional grammar, computational linguistics and corpus linguistics lies in the potential offered by each of the sub-disciplines aforementioned. Big Data can be readily accessed through data mining techniques, and the resulting output is in turn rapidly processed by corpus linguistics tools, ending up with a more refined inspection of the discursive tendencies observed in the target population due to the consultation of Halliday and Matthiessen’s (2014) full-fledged taxonomy. This blend may encourage future research and collaboration among researchers pertaining to such fields, should the outcome of this paper be proven feasible. The required steps to conduct this research and the overall planning of the workflow on Twitter’s data analysis is shown in Figure 4.

Step	Process	Description
1	Crawling	Extracting tweets and their metadata for posterior analysis and inspection.
2	Annotation	Adding or removing additional information referring to the data obtained.
3	Analysis	Heeding to statistical measures (collocations, frequencies, Word lists) for a richer understanding of the communicative event
4	Visualisation	Plotting visual aids which facilitate the interpretation of the data set and results obtained.

Figure 4: Structure of a Twitter’s research plan. Adapted from Burghardt (2015: 83).

Contrary to traditional sentiment-based analysis, this research shall neglect metadata such as user’s ID, tags, hyperlinks, and social interactions amongst tweets/retweets, since only linguistic cues (agents and actions) are deemed of relevance for the purposes of the present study. As for the tools used, Chorus Analytics shall gather raw data; SPSS is used to remove unnecessary information listed in the metadata, and to plot illustrative graphs; whereas AntConc is in charge of the actual linguistic analysis with collocations and frequencies. The software of choice employed for every step shall be explained and justified hereafter.

3.1. Chorus Analytics: Crawling and sentiment analysis

Chorus Analytics taps into Twitter’s Search API to conduct its analyses through two distinct modes: “the collection and visualisation suite” (Brooker *et al.*, 2016: 3). For the purposes of the present research, however, only the data-

gathering option is consulted, namely the Chorus TCDE (Tweet Catcher Desktop Edition). This interface enables the extraction of tweets by entering the search terms selected as the target object of study, as they are *Article 13* and *SaveYourInternet* in this particular case. Therefore, this software provides a first overview on the distribution of affective components (sentiment) throughout the preceding and following days of Article 13's approval, indicating both the polarities observed per day and the number of occurrences thereof. The period contemplated for tweet extraction spans from the 17th until the 28th of March 2019.

3.2. AntConc: Linguistic analysis

As noted by previous research, sentiment analysis software may be susceptible to nuances in the context, sentence structure (e.g. double negation) to interpret the output to the fullest extent (Taboada, 2016: 33). For this reason, the second part of this research employs the concordancing tool AntConc for a manual inspection of the most recurring patterns in the discursive construction of the selected event. After scrutinising the word list of the most frequently occurring lexical items (and filtering out those which bear no relation to the topic at hand, such as function words, references to hyperlinks, user IDs, retweets, etc.), the analysis shall proceed to spot the collocations for actions and agents and, whenever appropriate, resort to frequencies for a more comprehensive account on the discursive tendencies examined.

3.3. SPSS: Annotation and visualisation

For a more efficient data handling and examination, SPSS assumes the role of reviewing the raw data collected and modifying (adding or removing) them accordingly, should it become necessary. Regarding visual and graphic representations of the data, SPSS exhibits high efficiency in plotting bar/line graphs, and also allows for a flexible customisation of the variables displayed. This software shall perform the adequate statistical measures whenever hypothesis testing is required.

4. Data analysis and discussion

4.1. Sentiment analysis

After consulting Chorus TCDE, a total of 9,913 tweets² including the search terms *Article 13* and *SaveYourInternet* were retrieved in the time period between 17th and 28th of March 2019. The results are illustrated in the bar graph below:

² The data set obtained gathers any content a Twitter user publishes, including, but not limited to, posting retweets.

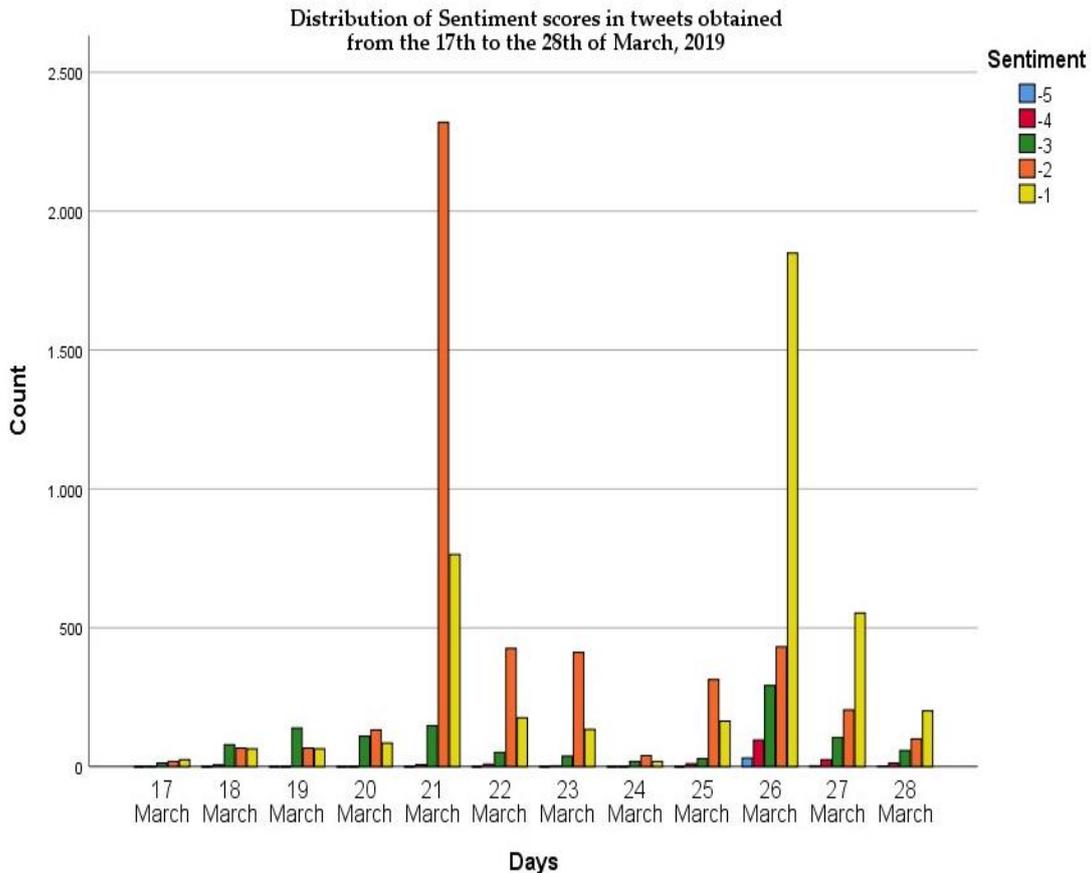


Figure 5: Bar graph on the distribution of sentiment scores in the tweets retrieved between 17th- 28th March 2019.

The first notable observation in Figure 5 is the lack of positive sentiments, but by no means does this imply the inexistence of positive viewpoints in the entire corpus. As a matter of fact, Chorus TCDE computes the total amount of both positive and negative sentiments. Afterwards, the researcher subtracts the difference between them, thus rendering a final score per tweet (sentiment score). Explicit tweets are quoted below³ to illustrate how the procedure works:

For all of you guys thinking that Brexit might have a chance to **remove** Article 13 from UK law, I've got **bad** news for you
[#SaveYourInternet](#) [#article13](#) [#brexit](#)
[#CopyrightDirective](#)

my god, most of the youtubers will **lose** their jobs as a youtuber... **thanks** to **fucking** article 13. world is **really fucked up** rn [#Article13](#)
[#SaveYourInternet](#)

Figure 6: Examples of sentiment scores' distribution in the tweets retrieved between 17th- 28th March 2019. Negative scores (-1) are highlighted in red, whereas positive items (+1) are highlighted in green.

³ Please note that the data are anonymised in order to protect the authors' identities.

As inferred from the instances shown in Figure 6, this software relies on "semantically driven" (Brooker *et al.*, 2016: 4) criteria to differentiate between positive and negative polarities. In the first instance, two negative items outweigh positive terms, thus rendering a final sentiment score of -1. In the second tweet shown above, intensifiers and verbs carrying negative connotations surpass positive lexis once again, only that the resulting sentiment score for the tweet becomes slightly more negative (-2).

Admittedly, categorising polarities does not entail restrictions on specific word classes (adjectives, verbal forms, nouns...), but considers negative connotations attached to the lexis itself. In fact, the algorithm employed by Chorus TCDE (Sentistrength) compiles a word list on positive/negative pairs based on human judgements offering features such as varying degrees of sentiment strength, spelling correction algorithm, and detection of emoticons, punctuation marks, repeated letters, intensifiers, etc. (Thelwall *et al.* 2010: 2552). Nevertheless, sentiment ratings may be misleading in instances with irony and/or sarcasm, like the "thanks" appearing in Figure 6. Due to this shortcoming, this paper balances its analysis with a manual inspection through AntConc.

Notwithstanding the apparent uneven distributions of sentiment scores shown in Figure 5, let us proceed to formulate the null hypothesis (H0: Sentiment scores act independently from the selected days) to test the starting hypothesis (H1: Sentiment scores will vary depending on the chosen day).

A chi-square test of independence was computed and found a statistically significant correlation between the categorical variables 'Sentiment scores' and 'Days' ($\chi^2(44, N = 9,913) = 3,368.767, p < .01$). The results appear compliant with the first observation above, as most of the negative sentiments encountered are represented on March 21 and 26, and therefore the null hypothesis is rejected.

	SENTIMENT SCORES					TOTAL COUNT
	-1	-2	-3	-4	-5	
17 March	25	19	13	1	0	58
18 March	64	67	79	6	0	216
19 March	64	67	139	0	0	270
20 March	85	132	109	0	0	326
21 March	764	2320	147	7	0	3,238
DAYS 22 March	176	426	51	9	0	662
23 March	134	411	38	3	0	586
24 March	19	40	19	0	0	78
25 March	164	313	29	10	0	516
26 March	1849	431	292	96	31	2,699
27 March	553	204	105	25	3	890
28 March	201	100	58	13	2	374
TOTAL COUNT	4,098	4,530	1,079	170	36	

Figure 7: Crosstabulation on Sentiment scores and Days.

The exact numbers and total count for each category are displayed in Figure 7 for a more accurate exploration. As far as the total count of items is concerned, March 21 is the most prolific day, followed by March 26 and 27. Concerning the sentiment types found, tweets seem to range preeminently from moderately negative (-2) to slightly negative (-1)⁴. The main distinction between the 21st and the 26th is that the former displays the highest tendency for moderate negativity (-2), whereas the latter gathers the 86.1% of the worst sentiment score in the list (-5), but also takes the highest percentage of the least negative stance (-1). This could lead to an interpretation of the 21st being mildly pessimistic about the uncertainty of the upcoming legal reform, while its approval on 26th March causes extreme reactions amongst Twitter users. Conversely, March 24 produces only a 0.5% of the total distribution of ‘-1’. As for ‘-5’, there are no extant records of it from the 17th to the 25th of March (0%).

The differing intensities in the production of negative sentiments seem to suggest a discursive pattern consisting of various stages: Preceding days to the event (17-20 March), Internet blackout day (21 March), transition between Internet blackout day and the approval of Article 13 (22-25), Article 13 approved (26), and the aftermath of the approval (27-28).

4.2. Linguistic analysis

Following the premise established in the sentiment-based analysis, a bar graph was plotted to illustrate the instantiations of process-types across the already mentioned group of days, as shown in Figure 8 below.

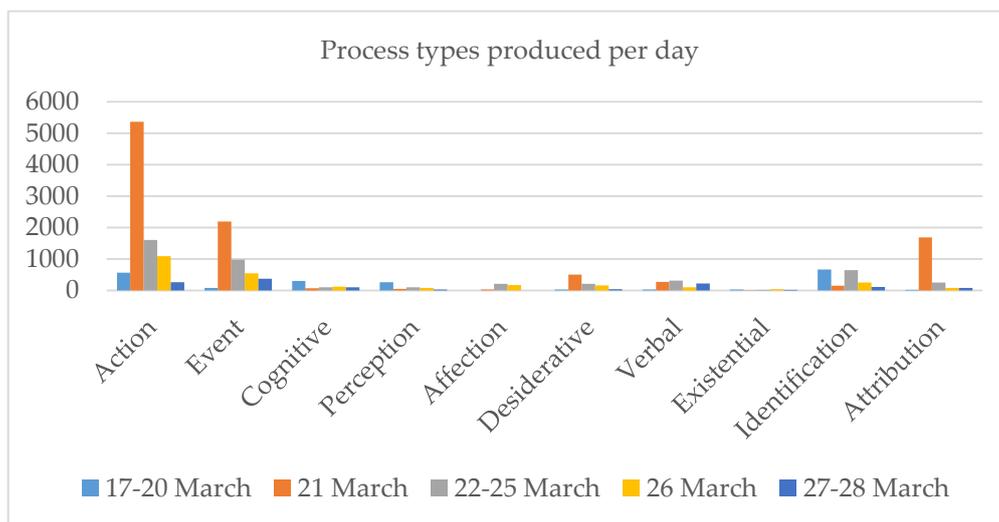


Figure 8: Process types found between the data-gathering period (17-28 March).

⁴ Please note that these labels refer to a scale representing overall sentiment strength (Thelwall et al., 2010: 2552) per tweet, ranging from slightly negative (-1) to extremely negative (-5).

The first noticeable aspect is the lack of behavioural and relational-possession cues throughout the whole time span⁵. As sharing traits, material processes outnumber the rest (especially action types), followed by relational types (identification and attribution) in the texts gathered across the time period specified. A pre-defined pattern in Figure 8 seems to be outlined, and yet some variations of the variables displayed make each day distinguishable.

During the first few days (17-20), the public's actions are centered upon relational identification, mostly collocating *Article 13* with *is* (e.g., "a survival battle", "the topic on the internet right now", "not a solution", "a worldwide issue"), also using attribution ("a bad reform", "a threat to collaboration and freedom"), and targeting those potentially affected by the reform ("If you *are* an EU citizen"). As anticipated, material action types are common at this stage, with verbs such as *vote* ("against Article 13"), *force* and *implement* ("Article 13 will force EU platforms to implement upload filters"), and *mark* ("Tomorrow marks an online demonstration against Article 13"). As for event sub-types, it is noteworthy to point at *fight*, *defend*, and *protest* being used intransitively, whilst *start* ("Action Week starts today") announces the collective event. Both mental perception and cognitive items could be used interchangeably in this text, as *see* ("many people fail to see the global trickle-down effect") and *know* ("most of you know Article 13") refer to comprehend the upcoming situation. As a remarkable trait, neither positive nor negative affection verbs are found. The only desiderative verb is *need* ("We need to stop Article 13"). Moving to existential cues, *there* precedes the sentence using *force* and *implement* mentioned above, thus reinforcing said statement ("There is no doubt"). As for the verbal category, the imperative *tell* directs the public's actions ("Tell your MEPs to delete Article 13").

The discourse seen on Internet Blackout day (21 March) seems proficient the most at material (both action and event) and relational-attribution processes. It introduces more diversity in the material-action domain with verbs such as *do* ("creators and viewers come together to do amazing things"), *limit* ("reform that could limit the content"), *undergo* ("Article 13 is going to undergo its final vote"), *block and upload* ("let a robot block the things you upload?"), *protest* ("Article 13"), *protect* ("the open internet"), *reach out* ("for your local MEPs") and *persuade*⁶ ("your MEP"). Event sub-types also add new actions like *live* ("If you live in the European Union") and *pass* ("If Article 13 passes"). As for relational-attribution, the same instances occur as in the previous day, but with higher frequencies, whereas identification sub-types drop in usage. The mental desiderative *need* assumes the same function as in the previous day, while *want*

⁵ For the sake of representativeness, any lexical item appearing 20 times or less has not been considered, unless otherwise specified.

⁶ Even though *persuade* could be conceived as a verbal process, these specific cases demand an impending action, hence its classification as a material process.

is added (“If you *want* to reach out”). Additionally, minor instances of affection processes appear (“Do not *worry*”), whereas perception types undergo no modification, except that they shrink in number. Likewise, existential cues remain the same, while verbal processes like *call* and *ask* co-occur with MEPs.

The next period comprised between 22 and 25 March is heavily influenced by material processes. Besides including the examples mentioned above, *oppose* (“Article 17, ex. Art. 13”) is added as an action, and *stand up* and *fail* as an event sub-type. Relational-identification processes shift from the use of *is* to *are* and *am* (“We *are* no bots”), often co-occurring with mental-affective types (“We *are* real people who *value* the freedom”), while the same applies to attributive types (“I *am* so *worried* about Article 13”). Opposed to the pragmatics of perception seen hitherto, *saw* is introduced in an agentless construction, which fulfills a time reference function (“Yesterday *saw* huge levels of online protest”). Even though their numbers do not increase substantially, instances of existential and desiderative types co-occur (“If *there* is anything I *wished* [...] I *want* the Article 13 to *fail*”). Lastly, *Contact* is introduced and classified as a verbal type, since the verbiage can be easily inferred (“*Contact* your MEPs to reject Article 13”).

On 26 March, more variety is shown in relation to material event types: (“liberty *died*”, “the EU parliament *voted* again today”, “Article 13 *has passed*”, “Article 13 *was approved*”), as well as in action processes (“*keep protesting* the article”, “The SaveYourInternet movement *has bravely fought* Article 13”, “*spread* awareness of Article 13”, “You *have ruined* the internet”). Mental affective components are diversified as well with *care* (“I actually *care* about Article 13”), *feel* (“I *feel* so bad”), and instantiations of swearing (“*fuck* everyone who approved Article 13”); the latter also co-occurring with desiderative sub-types (“*hope* you die”). Paradoxically, *hope* is barely used as a projection towards the future (“*hope* they will reject Article 13”), but partakes in swearing more frequently. Cognitive and perception processes do not undergo significant variation. Relational-identification’s lexis becomes drastically negative (“Article 13 *is* diabolical”, “Article 13 *is* complete bullshit”), even with rhetorical questioning (“*is* this the end?”), whereas attribution relates mainly to cursing (“whoever voted for Article 13 *is* an idiot”). On the other hand, the verbal form *explaining* alludes to an informative clip about the reform and its consequences. Other minor instances in this domain refer to *ask* as a sign of progressing towards a negotiation (“We *ask* you to reject the text of the directive”). As a final remark, existential cues display more variation: (“*there* is always a way”, “*there* is still hope”, “Article 13 vote *is happening* today”).

The production of processes decreases sharply in the ensuing days (27-28 March). A few additions to the repertoire include relational identification (“Article 13 *is* a step back to the past”), attributive (“Article 13 *is* not funny”), and existential types (“*there* are many unanswered questions”, “why is *there* no article 13 rebellion?”).

5. Discussion

In a broad sense, the “sentiment timeline” (Pang and Lee, 2008: 49) analysis may have hinted at the most prolific days in terms of issuing judgements and opinions on the subject matter. However, sentiment's usage does not seem to be entirely dependent on process types. As the linguistic analysis suggested, the highest sentiment production point in time (26 March) was not exceedingly dissimilar from other days when it comes to processes. From a qualitative perspective, higher negative scores in 26 March could be explained on the basis of the introduction of swearwords and other items such as attributes and enriched negative nouns, thus resulting in extreme polarities.

Despite displaying core features throughout the entire period examined (a seemingly coherent process type distribution), specificities in the pragmatics of the verbs observed outline a progression in the discourse that sets the cluster of days apart from each other, much in line with Socher *et al.*'s (2013: 1633) idea of widening the scope of analysis to interpret the semantics of the text adequately. For example, verbs like *vote* are employed either transitively or intransitively up until the 26th, when *vote* is nominalised (1,043 occurrences) since it becomes the topic/subject of discussion, and the public no longer demands an action⁷. Leaving aside verbs alluding to temporal references, process types also become stylistic features which seem to establish subtle boundaries between these texts/days, with verbal processes occurring on the periphery of 26 March as a sign of protest and/or request, the abundance of material processes preceding 26 March (referring to the Action Week's mobilisation) and their drop in usage afterwards, the prevalence of negative mental-affective processes during 26 March and a few days preceding it, which illustrates the public's discomfort with the resolution, etc.

Moreover, agency also distinguishes among the parts of the whole social event, as the collective *we* appears with a higher frequency right before the 26th of March, whereas the generic *creators and viewers* is maintained throughout every single day. As for the opposing parties, *Article 13* is the subject of the clause for the most part, while *EU parliament* fulfills the role of subject in a few instances only during the 26th of March. Lastly, *MEPs* is typically assuming the function of direct object, on the receiving end of verbal process types, but this trend is halted after the 26th of March.

6. Conclusions

To conclude, the use of sentiment scores in this study has been an effective guiding measure at the initial phase of the investigation to detect general trends and discursive patterns related to online activities across time, besides informing about the overall affiliation/disaffiliation with the subject

⁷ Even when there are instances of (in)transitive actions of this kind after the 26 March, these do not amount to representativeness if compared with their nominalised counterparts.

matter. It could be argued that, while the combination of sentiment and processes does not constitute a distinct genre altogether in the texts examined, their prominence at different points do account for sub-types, or rather stages within the whole social communicative event. As explored already in the SaveYourInternet movement, the discursive genres are adapted according to the pragmatic force which stands out the most at the stages examined: the introductory descriptive phase, followed by commanding with a great number of verbs in their imperative form, and a final turn towards expressive linguistic content.

On the other hand, despite the fact that the legal reform's wording might be modified in the future, the observed public commotion seems justified inasmuch as the ambiguous guidelines do not put forward a well-defined protocol. Instead, it seems that the conditions and penalties contemplated for publishing unauthorised work may be subject to the directive of each content sharing service provider, hence the uncertainty expressed by content creators and internet users in general. This uncertainty is reflected in the lexical choices the users made through Twitter posts and, whenever possible, further research should bridge the gap between legalese and lay communication with the purpose of providing a more comprehensive account on the stances of the parties involved. Albeit non-conventional, combining sentiment analysis with corpus linguistics and systemic functional linguistics has resulted in a fruitful approach to discourse and social practices, and thus collaboration in the respective fields is naturally encouraged.

Bibliography

- AGARWAL, Apoorv, Boyi XIE, Illia VOVSHA, Owen RAMBOW, and Rebecca PASSONNEAU (2011): "Sentiment analysis of Twitter data", in Meenakshi Nagarajan and Gamon, Michael (eds.): *Proceedings of the Workshop on Language in Social Media*. Portland: Association for Computational Linguistics, 30-38.
- BROOKER, Phillip, Julie BARNETT, and Timothy CRIBBIN (2016): "Doing social media analytics". *Big Data & Society*, 3(2), 1-12.
- BURGHARDT, Manuel (2015): "Introduction to tools and methods for the analysis of Twitter data". *10plus1: Living Linguistics*, 1(1), 74-91.
- CARDIE, Claire, Cynthia FARINA, Thomas BRUCE, and Erica WAGNER (2006): "Using natural language processing to improve eRulemaking", in José Fortes and Macintosh, Ann (eds.): *Proceedings of Digital Government Research*. San Diego, California: Digital Government Society of North America, 177-178.
- CHIKERSAL, Perna, Soujanya PORIA, Erick CAMBRIA, Alexander GELBUKH, and Chng Eng SIONG (2015): "Modelling Public Sentiment in Twitter: Using Linguistic Patterns to Enhance Supervised Learning", in Gelbukh, Alexander (ed.): *16th International Conference CICLing*. Cairo, Egypt: Association for Computational Linguistics, 49-65.
- HALLIDAY, Michael (1970): "Language Structure and Language Function", in John Lyons (ed.): *New Horizons in Linguistics*. Harmondsworth: Penguin, 213-229.
- HALLIDAY, Michael (1985): *An Introduction to Functional Grammar*. London: Edward Arnold.
- HALLIDAY, Michael and Christian MATTHIESSEN (2014): *Halliday's Introduction to Functional Grammar* (4th ed.). London: Routledge.
- KWON, Namhee, Stuart SHULMAN, and Eduard HOVY (2006): "Multidimensional text analysis for eRulemaking", in José Fortes and Macintosh, Ann (eds.): *Proceedings of Digital Government Research*. San Diego, California: Digital Government Society of North America, 157-166.
- MADIEGA, Tabiama André (2019): *Copyright in the digital single market 4th Ed.* [Online], available at [http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593564/EPRS_BRI\(2016\)593564_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593564/EPRS_BRI(2016)593564_EN.pdf) [consulted in March 2019].
- PANG, Bo and Lillian LEE (2008): "Opinion Mining and Sentiment Analysis". *Foundations and Trends in Information Retrieval*, 2(1), 1-135.
- SOCHER, Richard, Alex PERELYGIN, Jean WU, Jason CHUANG, Christopher MANNING, Andrew NG, and Cristopher POTTS (2013): "Recursive deep models for semantic compositionality over a sentiment treebank", in David Yarowsky, Baldwin, Timothy; Korhonen, Anna; Livescu, Karen; and Bethard, Steven (eds.): *Proceedings of the Conference on Empirical*

Methods in Natural Language Processing (EMNLP 2013). Melbourne, Australia: Association for Computational Linguistics, 1631-1642.

TABOADA, Maite (2016): "Sentiment analysis: An overview from linguistics". *Annual Review of Linguistics*, 2(1), 1-44.

THELWALL, Mike, Kevan BUCKETT, Georgios PALTOGLOU, and Di CAI (2010): "Sentiment strength detection in short informal text". *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

Fecha de recepción: 06/04/2019

Fecha de aceptación: 09/09/2019