Detecting Traces of Narrative Evolution on Telegram: Inductive Methods from Corpus-Based Discourse Analysis

Tom Willaert[1]

[1]Brussels School of Governance, IMEC-SMIT-VUB, Vrije Universiteit Brussel, Brussels, Belgium

Abstract

In the face of world-changing events, narratives on the messaging platform Telegram, including instances of disinformation, tend to arise and evolve at high speeds. However, key signals of this process, including newly emerging or idiosyncratic concepts, often elude traditional, top-down analyses. Addressing the need for inductive approaches to narrative evolution on Telegram, this paper operationalizes quantitative methods from the field of corpus-based discourse analysis. On a technical and methodological level, the paper discusses how data from Telegram's messages and images can be collected and preprocessed for the purposes of a 'keyness' (Log Ratio) analysis that surfaces salient nouns and verbs for further investigation. On an empirical level, this method is then applied to a case study of 225 predominantly Dutch-speaking Telegram channels (spanning the period March 2017- March 2022), revealing some of the dynamics that govern their recent shift from propagating narratives about the coronavirus pandemic to narratives concerning the war in Ukraine. This case study is accompanied by an interactive demonstrator that enables readers to further explore the processed dataset. The paper concludes with a reflection on the status of and future avenues for this 'distant reading' approach in relation to established interpretative practices.

1. Introduction

In political science, the concept of 'narrative' has broadly been defined as a form of discourse in which humans "construct disparate facts in [their] own worlds and weave them together cognitively in order to make sense of [their] reality" [1, p.135]. At a time when world-changing events such as pandemics and wars happen in rapid succession, this process of narrative sense-making is intensified on social media. There, spanning countless posts and channels, eclectic facts are continuously (re)combined into new stories, including instances of disinformation and conspiracy theory. A prototypical example of this are the narratives that circulate on Telegram, a messaging platform that through a lack of centralized content moderation tends to harbor conspiracy theories and other misleading or antagonistic discourse usually not tolerated on social media such as Twitter [2, 3]. In this prolific environment, newly-coined and often idiosyncratic concepts (such as the provocative 'denazification' used by the Russian government to legitimize the war in Ukraine) can emerge and propagate freely, which renders keyword-based query designs and other top-down methods for identifying narratives on the platform rather ineffective.

Confronting these challenges of (dis)information overload on Telegram and beyond, the development of inductive, machine-guided methods for mining narratives from (social media) texts at 'big data' scale has become an active area of research. First examples of such computational analyses of narratives can be traced back to work on scripts, story grammars, and planning formalisms from the field of artificial intelligence [4]. More recent continuations of this line of research have mapped the underlying structures and dynamics of narratives by representing them as (evolving) networks of relations between 'actants' figuring in texts, the latter concerning people, places, or organizations that are detected through techniques such as Named Entity Recognition (NER) [5]. Following a similar logic, some texts have explored the possibilities afforded by co-occurence networks of inductively-sourced hashtags to trace dynamics of converging narratives [6]. These empirically-informed approaches have thus yielded first insights into the structural ties that allow online conspiracy theories and other narratives to form from seemingly disparate concepts and information.

As this paper aims to elaborate, the study of online narratives, including the aforementioned network-based approaches, can benefit from bottom-up methods for identifying the idiosyncratic and evolving concepts that constitute those narratives. Previous literature in media studies has for instance bridged gaps between cultural-theoretical and computational-linguistic approaches by using

IJCAI 2022: Workshop on semantic techniques for narrative-based understanding, July 24, 2022, Vienna, Austria

tom.willaert@vub.be (T. Willaert)

word embeddings to demonstrate how platforms such as 4chan form incubators for "robust vernacular innovations" [7]. These conceptual innovations include neologisms such as 'redpill' (referring to an awakening from ignorance), whose emerging and shifting meanings can be interpreted as traces of the evolving narratives that define antagonistic subcultural communities. Comparable assumptions underpin quantitative, corpus-based analyses that trace the propagation of specific 'vernacular' concepts between platforms as means of identifying the 'mainstreaming' of fringe narratives [8, 9, 10]. Here, innovative or marked words that appear outside of the fringe environments in which they were first observed can be considered traces of a wider adoption of certain narratives.

Addressing the need for inductive approaches to narrative evolution on Telegram, the present paper operationalizes quantitative methods from the field of corpus-based discourse analysis. On a technical and methodological level, the paper offers a discussion of how data from messages and images from the platform can be collected and preprocessed for the purposes of a 'keyness' (Log Ratio) analysis that surfaces salient nouns and verbs for further investigation. On an empirical level, this method is then applied to a case study of 225 predominantly Dutch-speaking Telegram channels (spanning the period March 2017- March 2022). This case study tests the double hypothesis that 1) around the time of the outbreak of the war in Ukraine, Telegram channels that previously spread disinformation narratives on the coronavirus pandemic embraced narratives about the war, and that 2) this shift might reveal aspects of the underlying mechanisms governing the evolution of disinformation on the platform. To foster further exploration, this case study is accompanied by an interactive demonstrator that allows users to search and plot words from the dataset by their keyness scores. The paper concludes with a wider reflection on the status of and future avenues for this 'distant reading' approach to narratives in relation to established interpretative practices.

2. Data Collection

The focus of this paper is on the analysis of narrative evolution in message texts and images from public Telegram channels pertaining to Dutch-speaking far-right and conspiratorial communities. Following the taxonomy of platform affordances proposed in Van Raemdonck and Pierson [11], Telegram channels can be considered to afford "directed and isolated n-to-many interactions", meaning that the content travels from one channel to its many followers, who can receive and forward the content to other channels, but not respond to it. As such, Telegram channels can effectively be considered "depositories" and "amplifiers" of narratives [11, p.3].

Identifying relevant Telegram channels from which to mine these narratives is a non-trivial matter, as channels can be scattered and difficult to identify by channel name alone. Therefore, channels were retrieved by means of an established 'snowballing' method for Telegram research described in Peeters and Willaert [12]. This method repurposes Telegram's affordance of message-forwarding between channels as a means of identifying related channels. It assumes that if one channel forwards a message from another channel, a meaningful connection or shared interest exists between both. Starting from a seed list of channels defined based on expert knowledge, the researcher can thus retrace these links to other channels, bringing into view a network of interconnected channels in a bottom-up way.

For the purposes of this paper, a tailor-made scraper based on Python's Selenium library was used to automate and scale-up this process.[1] The network of channels under investigation in this article was first mapped in the summer of 2021. At that time, these channels were mainly preoccupied with the coronavirus pandemic and associated narratives, making this a suitable sample for exploring further narrative evolutions. The contents of these channels (both texts and images) were subsequently scraped again in March 2022. This results in a dataset of 821,020 messages from 225 public Telegram channels pertaining to Dutch-speaking far-right and conspiracy-theory communities, spanning a period between 18 March 2017 and 11 March 2022.

An initial inspection of this dataset revealed that narratives were constructed in both messages and images, with some images containing relevant patches of text. Working towards a 'multimodal' analysis that considers these aspects of the data, the channel contents were processed further along two tracks. Firstly, the texts embedded in the images were programmatically extracted using Google's Tesseract-OCR engine, by means of the Python-tesseract wrapper.[2] Secondly, the languages of the retrieved texts (from both posts and images) were detected using the Python 'langdetect' library.[3] This created opportunities for working

[1] https://selenium-python.readthedocs.io/
[2] https://pypi.org/project/pytesseract/
[3] https://pypi.org/project/langdetect/

with linguistically-homogeneous subcorpora in the subsequent analysis.

After preprocessing, it was found that of the retrieved messages, *ca.* 85% (697,364 messages) contained a non-empty message text field, and *ca.* 33% (267,956 messages) contained an image file.[4] After cleaning the outputs of the OCR for images, such as removing 'texts' that only contained new-line characters, texts could be extracted from *ca.* 67% (179,904) of the images. Automated language detection revealed that the corpus was multilingual (in part due to the forwarding of messages from international channels), with English and Dutch being the most prominent languages. Of the message texts, *ca.* 21% (143,120) were classified as written in English, and *ca.* 57% (399,842) in Dutch. For the images from which texts were extracted, we found *ca.* 53% (94,803) contained text in English, and *ca.* 28% (51,138) contained text in Dutch. The prominence of English texts in the images again points towards an international dynamic of message and content forwarding between channels.

3. Methodology

In order to inductively detect signals of narrative evolution in the collected data, this paper applies the method of 'keyness' analysis. This approach from the fields of corpus linguistics and corpus-based discourse analysis is directed at identifying 'key' items (e.g. words) in a target corpus in relation to a reference corpus based on the frequencies of items in both corpora. As such, a keyness analysis can support an exploratory approach to texts that gives an indication of their "aboutness" [13, p.227]. Arguably, this makes the method well suited for our purposes of identifying emerging narrative signals in texts. The keyness metric chosen for this paper is that of Log Ratio, which is defined as the "binary log of the ratio of relative frequencies" [14]. This gives a measure of the actual observed difference between two corpora for a key item (rather than a measure of statistical significance). The advantage of this is that it allows for the sorting of items by the size of the actual frequency difference between the corpora, enabling us to find the top N most key items. In order to calculate the Log Ratio for an item in target corpus C1 and a reference corpus C2, we take the binary logarithm of the ratio of the normalised frequencies of the term in C1 and C2 (which are each multiplied by a factor of 1,000,000 for readability purposes).[5]

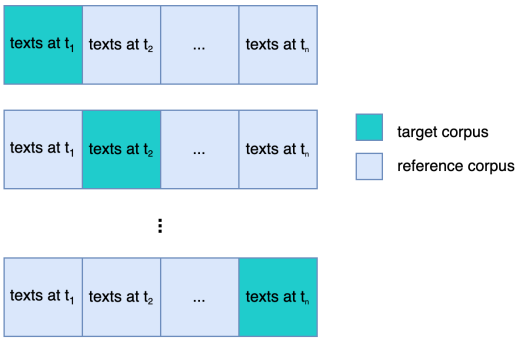

Figure 1: Schematic overview of the approach. Data are grouped by timestamps. For data at each timestamp (the target corpus), keyness scores (Log Ratio) for nouns and verbs are calculated in relation to data for all remaining timestamps (the reference corpus). Offering a 'distant' perspective of the period as a whole, this approach foregrounds key items for each individual week in relation to the full corpus minus that week.

As illustrated in Figure 1, our overall approach to narrative detection on Telegram, then, is to detect these key items from a reference corpus of texts grouped by week in relation to all remaining data. We then consider the items with the highest keyness scores for each timestamp, thus opening them up for further interpretation. Concretely, this technical pipeline comprises the following steps:

1. We filter the data by content type (message texts, image texts, or combinations of both) and language (Dutch or English).
2. We group the texts by timestamps (viz. per week of data).
3. We clean the texts at each timestamp by removing hyperlinks and emojis.
4. We perform part of speech tagging and retain only nouns and verbs (as we consider these to express core concepts).
5. We calculate the frequencies for these items per timestamp (week).
6. We calculate the Log Ratio of the target corpus (normalized frequencies) in relation to all other weeks.
7. Finally, we rank words by keyness score.

On a conceptual level, this approach returns keyness scores for items in relation to the combined

[4] It should be acknowledged here that for messages with multiple images, the scraper only stored the first image attached to the message.

[5] For a Python implementation, see https://kristopherkyle.github.io/corpus-analysis-python/Python_Tutorial_7.html

data that precede and follow it – offering a distant perspective on distinctive (key) narrative signals for each week's worth of data in relation to the full period minus that week. The keyness scores for the final timestamp have a special status in this regard, as they reveal key items in relation to all of the preceding data, illustrating what is key at the last moment of observation. It should be acknowledged upfront that this keyness analysis does not yet integrate semantics, apart from the significance attributed to nouns and verbs as key indicators of narratives. As will be expanded upon in the conclusion, this approach thus requires further interpretation and contextualization of the detected key items.

In order to illustrate this method and make an empirical contribution to the study of narrative dynamics on Telegram, the following section zooms in on a case study that investigates the relation between narratives about the coronavirus pandemic and the war in Ukraine as expressed in our corpus.

4. Case Study and Findings

Recent and on-going events such as the coronavirus pandemic and the war in Ukraine have kindled an interest in the evolutionary dynamics of (disinformation) narratives among researchers, civil society actors, and journalists. One comparative analysis of international fact-checks has for instance revealed some striking, high-level parallels between disinformation surrounding both events in terms of style and contents [15]. Examples from this study include references to Nazism (e.g. the coronapass as a Nazi 'health passport' or Ukraine as a region that should be 'denazified'), and recurring conspiracies about secret laboratories (e.g. false claims that the coronavirus was created in a lab and references to the alleged presence of U.S. bioweapon labratories in Ukraine as a pretext for the war). This then raises the question of whether similar trends are reflected on a more localized level. Or more concretely: have the same communities that previously pushed false narratives about the coronavirus also embraced disinformation about the war in Ukraine?

A recent study by the Institute of Strategic Dialogue confirms that this can indeed be the case [16]. Based on the analysis of a dataset of 229 German-language Telegram channels (spanning the period between 1 November 2021 and 27 February 2022) pertaining to far-right and conspiracy theory communities, this study has shown that terms from a preconstructed list of 80 keywords related to "Russia, Ukraine, the breakaway regions in Eastern Ukraine and the Russia-Ukraine crisis in general" indeed become more frequent in the discourse of these communities. The actual (pro-Russian) narratives themselves were then analysed on the basis of close-readings of articles from the most frequently shared domains in the dataset [idem.]. As the dataset investigated in the aforementioned study closely resembles the one introduced in the present article, we can hypothesize that a similar transition from coronavirus-related narratives to narratives about the war in Ukraine should be observable in our corpus. Moreover, we can also hypothesize that our inductive approach can reveal more detailed traces of the actual narratives that thus emerge, thus opening up perspectives on the more fundamental dynamics underlying this narrative evolution.

In order to interpret the results of the keyness analysis in light of these hypotheses, they have been integrated into an interactive demonstrator or 'observatory' [17] that allows for interactive exploration and plotting of terms based on their keyness scores. This 'observatory' covers the full dataset (only snapshots of which are discussed in the present paper) and is openly available online.[6]

A first observation that can be made on the basis of our keyness analysis, is that we can indeed see emerging traces of narratives concerning the war in Ukraine. The table in Figure 2 shows the top 20 nouns and verbs (by keyness score) retrieved for the last four weeks of English message texts in the dataset. From this overview, it follows that discourse in these messages distinguishes itself from previous weeks through references to the war in Ukraine. Possible first traces are already observed in the week of February 20 in the form of a reference to "mobilisation". Further, more explicit references can be found in ensuing weeks, which feature high-keyness words such as "demilitarize" and "bombards" (week of 27/02/2022), "defections" (week of 06/03/2022), as well as "vladimir" and "corridors" (week of 13/03/2022).

A second observation is that our empirical analysis reflects some of the trends observed in the aforementioned study of narrative similarities in fact-checks. Among the high-keyness terms that are detected in the latter weeks of the dataset, terms legitimizing the war such as "denazify" (27/02/2022) clearly evoke Nazism. The analysis likewise foregrounds references to the biolaboratories conspiracy mentioned earlier (e.g. "biolaboratories", "bioscientist" (13/03/2022)). Results of a wider search for terms referring to biology laboratories shown in Figure 3 reveal that an earlier segment of the

[6] https://jvansoest.github.io/

2022-02-20	2022-02-27	2022-03-06	2022-03-13
pooling: 9.56	demilitarize: 9.26	actuaries: 8.37	biolaboratories: 8.74
admires: 8.75	sotu: 8.52	rearranging: 8.37	cleanups: 7.87
cma: 8.34	bombards: 8.35	underestimating: 8.37	consciences: 7.87
interferon: 8.34	denazify: 8.16	modernize: 8.05	gf: 7.87
methylation: 8.34	aggravate: 7.93	defections: 7.64	hyperloops: 7.87
symptomstake: 8.34	contaminations: 7.35	dispelling: 7.64	liquidation: 7.87
ínside: 8.34	extrapolated: 7.35	gloat: 7.64	optic: 7.87
canberra: 7.75	geïnfiltreerd: 7.35	plucked: 7.64	shekels: 7.87
christened: 7.75	gibberish: 7.35	thalidomide: 7.64	wheather: 7.87
cores: 7.75	holdthelove: 7.35	archeology: 7.05	releqsed: 7.61
disheartened: 7.75	junta: 7.35	boodschappen: 7.05	corridors: 7.29
dubai: 7.75	locators: 7.35	centralbanks: 7.05	hotlines: 7.29
hersenen: 7.75	occupies: 7.35	conjecture: 7.05	vladimir: 7.29
implanteren: 7.75	ores: 7.35	cyberaanvallen: 7.05	airfields: 6.87
lrads: 7.75	peacefull: 7.35	debtsystem: 7.05	asholes: 6.87
mobilisation: 7.75	perfected: 7.35	depts: 7.05	batting: 6.87
multistakeholder: 7.75	pounding: 7.35	desiring: 7.05	bioscientist: 6.87
reagents: 7.75	regio: 7.35	devoured: 7.05	boostered: 6.87
specimens: 7.75	renember: 7.35	digitisation: 7.05	buttercups: 6.87
spliced: 7.75	slut: 7.35	doodeng: 7.05	carefulled: 6.87

Figure 2: Top 20 nouns and verbs with highest keyness scores for message texts in English for the last four weeks of the dataset (week of 20/02/2022 - week of 13/03/2022). Various traces of emerging narratives about the war in Ukraine can be observed (e.g. "mobilisation", "demilitarize", "denazify", "vladimir"). This indicates that the same Telegram channels known for propagating narratives about the coronavirus pandemic have recently also embraced narratives about the war in Ukraine

data where this term had a higher score was during the coronavirus pandemic, before the Ukraine war. This illustrates that some narrative traces are actually recurrent in the dataset. Moreover, this plot demonstrates that the dataset contains a relatively stable narrative 'undercurrent' marked by e.g. words referring to the coronavirus pandemic. These have a keyness score that remains close to 0 in each week of the dataset.

Finally, the repeated occurrence of 'biolaboratories' as a high-keyness item suggests that within the Dutch-speaking disinformation communities under investigation, narratives contextualizing the war in Ukraine are but a salient pivot point in an ongoing process of narrative evolution. Offering a more 'zoomed out' perspective, Figure 4 shows the results of the keyness analysis for texts in English from both messages and images combined, covering a more extended period of time. A closer inspection of key items predating the Russian invasion of Ukraine hint at a range of other events that have been appropriated to match the then-predominant agenda of the communities. Our method for instance picks up traces of references to the freedom convoys in Canada (e.g. "Winnipeg", "blockading" (06/02/2022)), which demonstrates how Telegram channels continuously adapt narratives to match ongoing events.

5. Discussion

In light of our hypotheses, the analysis conducted above indeed reveals traces of narratives related to the war in Ukraine in communities that were previously mainly concerned with the pandemic. Furthermore, our inductive approach brings into view three more general dynamics governing this transition. Firstly, it was possible to observe both emerging narratives as well as more 'stable' undercurrents. Secondly, our case study suggests that certain narratives recur over time. Thirdly, expanding the scope of the investigation indicates that the recent shifts between narratives are part of a longer process of narrative evolution.

Given the specific nature of the corpus under consideration, these observations might provide some deeper insights into the nature of disinformation narratives. It notably seems to be the case that in order to persist, disinformation needs to contain a foundation of recognisable, recurring elements, yet at the same time it needs to be flexible enough to adapt to world-changing events. It can be argued that on Telegram, this continuous process of recurrence and adaptation is facilitated by the permissive affordances of the platform.

6. Conclusions and Future Work

This paper set out to make a double contribution to the detection of evolving, often idiosyncratic narratives on social media. For one thing, the paper proposed a technical pipeline for detecting traces of narrative innovations and narrative continuity in a bottom-up way by operationalizing keyness analysis (Log Ratio). For another, the paper applied this method to the case study of narrative evolution on Dutch-speaking Telegram channels (pertaining to far-right and conspiracy theory communities). It has thus been shown how keyness analysis can be applied to Telegram data to inductively identify traces of emerging or persistent narratives that might warrant further investigation.

It should be acknowledged that the exploratory scope of the present paper has its limitations. Building on these initial results, at least two pathways for future research can be envisaged. On a methodological and technical level, more work is needed to reduce noise and introduce additional granularity in the analysis. As has been illustrated in this

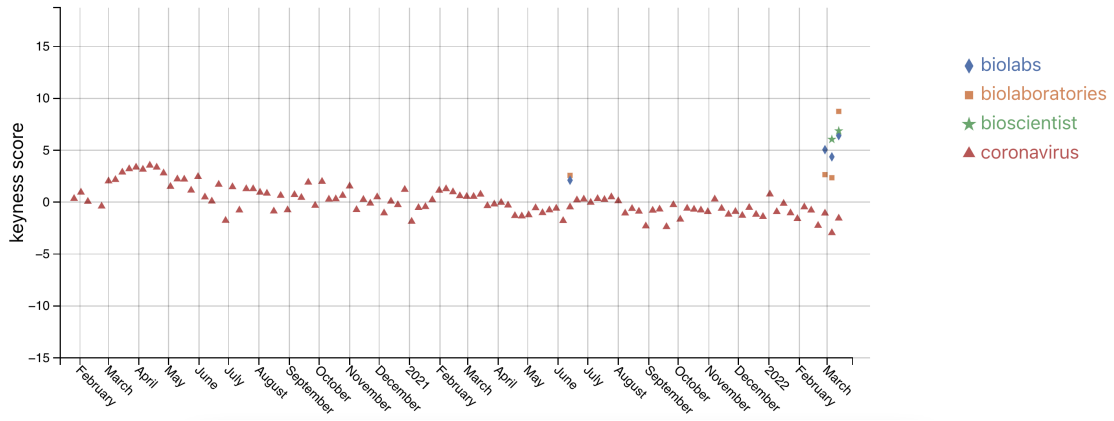

Figure 3: Plot of keyness scores over time for terms referring to biology laboratories and the coronavirus in the dataset's message texts in English. The graph suggests a recurrence of emerging narratives involving biology laboratories during the pandemic and at the start of the war in Ukraine. The keyness score of the term "coronavirus" remains close to 0 in each week of the dataset (except for some higher scores around the time of the outbreak of the pandemic), suggesting a relatively stable 'undercurrent' of coronavirus-related narratives

2022-01-30	2022-02-06	2022-02-13	2022-02-20	2022-02-27	2022-03-06	2022-03-13
plexiglass: 9.15	gargle: 9.79	inclusions: 9.17	pooling: 9.48	sotu: 8.50	actuaries: 8.32	biolaboratories: 8.44
país: 8.42	bbq: 9.21	gcs: 8.85	admires: 8.67	bombards: 8.33	rearranging: 8.32	biomaterials: 8.33
alsjeblieft: 7.83	gargling: 9.21	erkent: 8.43	interferon: 8.26	aggravate: 7.92	modernize: 8.00	gf: 8.33
auis: 7.83	winnipeg: 9.21	schrapt: 8.43	methylation: 8.26	integrates: 7.92	defections: 7.58	hereinafter: 8.33
cabinets: 7.83	embalmers: 8.21	unknown: 8.07	plumbers: 8.26	lich: 7.92	dispelling: 7.58	igi: 8.33
cama: 7.83	irrationality: 8.21	cardiologists: 7.85	symptomstake: 8.26	rcp: 7.92	excised: 7.58	ocr: 8.33
cohoots: 7.83	weathers: 8.21	cooperations: 7.85	ínside: 8.26	shin: 7.92	furor: 7.58	tena: 8.33
elmo: 7.83	articulates: 7.62	debuted: 7.85	canberra: 7.67	tamara: 7.92	gloat: 7.58	beau: 7.74
flyby: 7.83	artis: 7.62	dragons: 7.85	chid: 7.67	demilitarize: 7.65	plucked: 7.58	cfs: 7.74
hoofdstad: 7.83	blockading: 7.62	epoch: 7.85	christened: 7.67	denazify: 7.50	ruble: 7.58	chu: 7.74
jede: 7.83	caverns: 7.62	firewood: 7.85	disheartened: 7.67	cleavage: 7.33	underestimating: 7.32	cleanups: 7.74
landverraad: 7.83	changers: 7.62	fling: 7.85	dubai: 7.67	contaminations: 7.33	archeology: 7.00	consciences: 7.74
lrem: 7.83	deletetheelite: 7.62	fran: 7.85	forcitizens: 7.67	extrapolated: 7.33	boodschappen: 7.00	dase: 7.74
maladministration: 7.83	destabilizes: 7.62	gloss: 7.85	hersenen: 7.67	furin: 7.33	cami: 7.00	dens: 7.74
malformations: 7.83	disseminators: 7.62	housekeeper: 7.85	imbecile: 7.67	geïnfiltreerd: 7.33	centralbanks: 7.00	emperors: 7.74
massed: 7.83	endosomes: 7.62	imprints: 7.85	implanteren: 7.67	gibberish: 7.33	cmos: 7.00	esos: 7.74
müssen: 7.83	enor: 7.62	inductor: 7.85	intation: 7.67	holdthelove: 7.33	conjecture: 7.00	gore: 7.74
obsessing: 7.83	gekker: 7.62	inscriptions: 7.85	lrads: 7.67	inchiding: 7.33	crocodile: 7.00	guarantor: 7.74
obwohl: 7.83	irate: 7.62	lastnight: 7.85	mobilisation: 7.67	junta: 7.33	cyberaanvallen: 7.00	herr: 7.74
orking: 7.83	kellie: 7.62	levee: 7.85	multistakeholder: 7.67	locators: 7.33	debtsystem: 7.00	historical: 7.74

Figure 4: Results of the keyness analysis for texts in English from both messages and images combined, covering a more extended period of time. A closer inspection of key items predating the Russian invasion of Ukraine hint at a range of other events that have been appropriated to match the then-predominant agenda of the retrieved channels, including the 'freedom convoys' in Canada ("Winnipeg", "blockading"). This points towards a continuous process of narrative evolution on Telegram

paper, transferring methods from corpus-based discourse analysis to Telegram requires intensive data-preprocessing. Future investigations might for instance explore more refined methods for language detection and optical character recognition (text extraction from images) suitable for Telegram's idiosyncratic (visual) discourse. Along the same lines, future research might complement the aggregated

perspective on offer and explore the distributions of key items over channels, thus bringing into perspective more intricate relations between channel dynamics and discourse. Finally, additional methodological work is needed to situate the retrieved items in their wider semantic networks, for instance through statistically-informed co-occurrence analyses. Introducing further granularity, one promising avenue here would be to contextualize key items through graph-like representations of narratives inferred from the sentences' argument structure [18].

On a more conceptual level, our analysis raises bigger questions of meaning and interpretation. As indicated, the keyness analysis itself does not capture the semantics of the messages and image texts under investigation. Meaning has to be assigned to key items by the human interpreter, for instance by considering and comparing combinations of key items, by looking up the retrieved key words in the corpus and reading the messages or image texts in which they figure, or through broader cultural or media-theoretical contextualization. This foregrounds the question of how critical frameworks might be developed that streamline and formalize the integration of inductive methods from data science and interpretative approaches from the humanities. Proposals for such frameworks have been made under the denominator of 'data hermeneutics' [19, 20], opening up the field for future work on actionable implementations.

7. Data availability statement

The dataset of weekly keyness scores for nouns and verbs in the Telegram dataset can be queried through the online demonstrator accompanying the paper.

Acknowledgments

This project has received funding from the European Union under Grant Agreement number INEA/CEF/ICT/A2020/2394296. The paper was written during a research visit at SciencesPo Médialab made possible by a travel grant from the Research Foundation Flanders (FWO). The author wishes to thank Jeroen Van Soest (Vrije Universiteit Brussel) for his work on the demonstrator accompanying this paper.

References

[1] M. Patterson, K. R. Monroe, Narrative in political science, Annual Review of Political Science 1 (1998) 315–331. doi:10.1146/annurev.polisci.1.1.315.

[2] R. Rogers, Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media, European Journal of Communication 35 (2020) 213–229. doi:10.1177/0267323120922066.

[3] A. Urman, S. Katz, What they do in the shadows: examining the far-right networks on Telegram, Information, Communication & Society (2020) 1–20. doi:10.1080/1369118X.2020.1803946.

[4] I. Mani, Computational narratology, in: P. Hühn, J. Pier, W. Schmid, J. Schönert (Eds.), The Living Handbook of Narratology, Hamburg University, 2013. URL: http://www.lhn.uni-hamburg.de/article/computational-narratology.

[5] T. R. Tangherlini, S. Shahsavari, B. Shahbazi, E. Ebrahimzadeh, V. Roychowdhury, An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate, Pizzagate and storytelling on the web, PLOS ONE 15 (2020) e0233879. doi:10.1371/journal.pone.0233879.

[6] M. Tuters, T. Willaert, Deep state phobia: Narrative convergence in coronavirus conspiracism on Instagram, Convergence: The International Journal of Research into New Media Technologies (in print).

[7] S. Peeters, M. Tuters, T. Willaert, D. de Zeeuw, On the vernacular language games of an antagonistic online subculture, Frontiers in Big Data 4 (2021) 1–15. doi:10.3389/fdata.2021.718368.

[8] S. Peeters, T. Willaert, M. Tuters, Travelling tokens: Following extreme terms from 4chan/pol/ to Breitbart. OILab blog, https://oilab.eu/travelling-tokens-following-extreme-terms-from-4chan-pol-to-breitbart/, 2020.

[9] T. Willaert, P. V. Eecke, J. V. Soest, K. Beuls, A tool for tracking the propagation of words on Reddit, Computational Communication Research 3 (2021) 117–132.

[10] S. Peeters, T. Willaert, M. Tuters, K. Beuls, P. Van Eecke, J. Van Soest, A fringe mainstreamed, or tracing antagonistic slang between 4chan and Breitbart before and after Trump, in: R. Rogers (Ed.), How Misinformation Propagates on Social Media. Mainstream-

ing the Fringe, Amsterdam University Press, in print.
[11] N. Van Raemdonck, J. Pierson, Taxonomy of social network platform affordances for group interactions, in: 2021 14th CMI International Conference - Critical ICT Infrastructures and Platforms (CMI), 2021, p. 1–8. doi:10.1109/CMI53512.2021.9663773.
[12] S. Peeters, T. Willaert, Telegram and digital methods: Mapping networked conspiracy theories through platform affordances, M/C Journal 25 (2022). URL: https://journal.media-culture.org.au/index.php/mcjournal/article/view/2878. doi:10.5204/mcj.2878.
[13] C. Gabrielatos, Keyness analysis, in: C. Taylor, A. Marchi (Eds.), Corpus Approaches To Discourse, Routledge, 2018, p. 225–258. URL: https://www.taylorfrancis.com/books/9781351716079/chapters/10.4324/9781315179346-11. doi:10.4324/9781315179346-11.
[14] A. Hardie, Log Ratio: an informal introduction. ESRC Centre for Corpus Approaches to Social Science (CASS), http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/, 2014. URL: http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/.
[15] T. Willaert, M. G. Sessa, From infodemic to information war: A contextualization of current narrative trends and evolutions in Dutch-language disinformation communities, https://researchportal.vub.be/en/publications/from-infodemic-to-information-war-edmo-belux-investigative-report, 2022.
[16] J. Smirnova, P. Matlach, F. Arcostanzo, Support from the conspiracy corner: German-language disinformation about the Russian invasion of Ukraine on Telegram, https://www.isdglobal.org/digital_dispatches/support-from-the-conspiracy-corner-german-language-disinformation-about-the-russian-invasion-of-ukraine-on-telegram/, 2022.
[17] T. Willaert, P. Van Eecke, K. Beuls, L. Steels, Building social media observatories for monitoring online opinion dynamics, Social Media + Society 6 (2020) 1–12. doi:10.1177/2056305119898778.
[18] E. Ash, G. Gauthier, P. Widmer, RELATIO: Text semantics capture political and economic narratives, arXiv (2021). URL: https://arxiv.org/abs/2108.01720. doi:10.48550/ARXIV.2108.01720.
[19] A. Romele, M. Severo, P. Furia, Digital hermeneutics: From interpreting with machines to interpretational machines, AI & SOCIETY 35 (2020) 73–86. doi:10.1007/s00146-018-0856-2.
[20] P. Gerbaudo, From data analytics to data hermeneutics: Online political discussions, digital methods and the continuing relevance of interpretive approaches, Digital Culture & Society 2 (2016) 95–112. doi:10.14361/dcs-2016-0207.