

QAnon: Authorship Attribution in a Group of Six Suspects

Stylometric analyses reveal who in QAnon's publication management group writes with a personal style closest to QAnon's Socratic style

OrphAnalytics' experts:

Stylometric analyses and project management: Claude-Alain Roten

Barycentric algorithm: Léonard Studer

Computer science: Antoine Jover

Critical evaluation: Marc Zehnder

Linguist: Constanze Andel

Evaluation from a criminological perspective: Elise Soulier

Evaluation from a writing analysis perspective: Sylviane Valley-Muehlethaler

Communication: Lionel Pousaz

contact@orphanalytics.com

www.orphanalytics.com

Executive Summary

In 2020, OrphAnalytics conducted his first series of stylometric analyses on the QAnon corpus – i.e. the whole set of online messages known as “Q-drops” attributed to an anonymous named “Q”. Published in December 2020 as a white paper, our first QAnon investigation revealed two different periods, namely Q1 (Oct. 28 to Dec. 1st, 2017) and Q2 (Dec.1, 2017, to Nov. 13, 2020), each associated with two different styles, as revealed by our stylometric analysis, as well as two different publishing platforms (Q1 on 4chan and Q2 on 8chan/8kun). Texts in both groups Q1 and Q2 are likely the product of two different authors or groups of authors.

The existence of two individual styles objectively contradicts the belief of QAnon supporters that a single author, presented as a senior U.S. official, is the sole author of the controversial messages that inspired the January 6, 2021 riots on Capitol Hill in Washington, D.C. But our initial analysis could not yet address the possible identity of the real person(s) behind QAnon messages, since we did not have enough texts from our potential suspects to compare with the Q-drops at the time.

Our potential candidates are six individuals heavily involved in disseminating QAnon publications from early on. They have been identified in several media investigations as possible authors: Paul Furber, Coleman Rogers, Tracey Diaz, Cristina Urso, Jim Watkins, and Ron Watkins. None of them have publicly admitted to participating in the writing of these posts (8chan's IT manager Ron Watkins could be considered an exception since he indirectly admitted participation in a video interview¹ before recanting).

Nowadays, a sufficient number of available texts can be safely attributed to these candidates. This new material allows us to run a second analysis to identify the likely QAnon author(s) by style

¹ Ron Watkin's interview at the 49th min. in the 6th of the 6-episode documentary *Into the Storm* (Hoback, 2021) of HBO directed by Cullen Hoback. See Appendix B.

comparison. Stylometric analyses allow us to measure style similarities, and thus assign an anonymous text to a candidate, or to decipher which individual from a given group writes with the style closest to a given anonymous text.

In order to measure which of the six candidates' styles are the closest from both Q1 and Q2 Q-drops, we used OrphAnalytics patented method, relying on the principle of Voronoi diagrams, and measured the distance between texts or groups of texts. For this purpose, we established a usage profile of trigrams (i.e. three characters chains) composing sequences. This allowed us to represent each sequence with a point in a multidimensional space where each dimension corresponds to the frequency of a specific trigram. The similarity measure between two sequences corresponds to the distance between the vector of each sequence and the average vector of the texts of a suspect.

This distance enabled us to associate one of our six candidates with the style of QAnon. It confirmed the gap already observed in our first white paper between the two periods/styles Q1 and Q2.

Our stylometric analysis by distance shows that Ron Watkins' texts are by far the closest to those of Q2, that is to say the vast majority of QAnon's corpus. At this stage, we are confident that among our six suspects, Ron Watkins is by very far the best match with the Q-drops published during the Q2 period. Moreover, it is important to keep in mind that he was the IT manager of the 8chan/8kun platform that published the entirety of the Q-drops during the same Q2 period. In addition, no more Q-drops have been published after he resigned as the 8kun IT manager in November 2020.

For the messages belonging to the Q1 period/style, our results are less clear-cut. It is even possible that the author of Q1 is not among the group of suspects. But our analysis seems rather to suggest that the style of Q1 is the result of a collaboration among several authors. Our stylometric analyses point to Paul Furber, Cristina Urso, and Ron Watkins.

Besides our stylometric analysis by distance, we note that QAnon messages are characterized by a striking feature: their high amount of interrogative sentences revealing what could be described as Socratic questioning. The corpus in Q1 has a higher question rate (53.3% number of characters in questions normalized by the total number of characters) than Q2 (31.6%). These question rates are much higher than those of the texts from our candidates used for comparison. Only the writings of Ron Watkins and of Paul Furber somewhat come close with hardly 11-12%.

Another style analysis shows that QAnon texts are the product of specific formatting: short sentences with complex terms. The suspects' texts differ, displaying a more usual pattern characterized either by short sentences with simple words or by long sentences with complex words. As a result, QAnon's formatted writing sounds artificial. While it is easy to reduce the size of sentences in a text, it is more difficult to enrich it with complex terms. Ron Watkins is the candidate whose style most closely matches QAnon's. His complex vocabulary combined with his relatively short sentences resembles QAnon's.

The candidates' writings that we use as comparison are of a different nature than the Q-drops and their constrained, idiosyncratic style. As such, we should not expect a perfect match with the personal writings from our candidates.

Finally, it is worth noting that Paul Furber, a likely co-author during the Q1 period according to our analyses, has publicly criticized a decline in Q-drops quality from the Q2 period, right after the transfer from 4chan to 8chan². His reaction could suggest the frustration of an ousted co-author.

The change of sites from 4chan to 8chan was associated with the will to secure the publication of QAnon by tripcodes. The transition from Q1 to Q2 seems to mark the transformation of the QAnon style marked by complex writing, cooperative at times, to a simpler style, after what might have been a takeover of QAnon publishing by Ron Watkins.

Introduction

Between late 2017 and late 2020, 4953 anonymous messages were written to create and control the QAnon movement in the United States. For sympathizers, these anonymous messages were written by one single person: a senior U.S. administration official with the "Q" clearance that would give him access to the U.S. military's nuclear secrets. These QAnon messages (Q-messages anonymously written) reinforce the mythology of the Deep State proclaiming that an elite secretly governs the United States. This elite would be composed of political executives (mainly members of the Democratic Party), business leaders and show business people, all allegedly involved in mass pedophile abuse. During his presidency (2016-2021), Donald Trump would have been fighting the Deep State.

Several suspects have been proposed as possible authors the QAnon corpus: Donald Trump, Steve Bannon for example (Hoback, 2021). But some media investigations have focused on the people involved in the publication and media coverage of the corpus (Zadrozny & Collins, 2018; Wang & Click, 2020).

Four people were involved in the 4chan period. South African Paul Furber (aka Baruch the Scribe) claimed to be the first to detect the first Q-drops (Furber, 2021). Furber was responsible for publishing Q-drops on the 4chan site. Coleman Rogers (aka Pamphlet Anon) helped manage QAnon's publishing during this period. Soon the two moderators received the collaboration of Tracy Diaz (aka Tracy Beanz) to effectively mediatize these messages. Finally, this group benefitted from the assistance of Rogers' wife, Cristina Urso (aka Radix Verum).

During the transition from 4chan to 8chan, two other people appeared in the management of the publication of Q-drops: the owner of the 8chan site Jim Watkins (aka seriousjim) and his son Ron (aka CodeMonkeyZ) who was the IT administrator of the site at that time and thus had access to the data of the user accounts.

After the presidential election won by Joe Biden, Donald Trump refused to publicly acknowledge his defeat. Trump's attitude reinforced a protest movement composed in part of QAnon supporters. Growing discontent with the election led to the Capitol Hill riots in Washington D.C., on January 6, 2021.

² Paul Furber's interview at 41st min. in the 3rd of the 6-episode documentary *Q : Into the Storm* (Hoback, 2021) of HBO directed by Cullen Hoback.

In response to growing tensions during the U.S. election, OrphAnalytics undertook stylometric analyses of the QAnon corpus in mid-October to determine the number of authors in the corpus, and furthermore to identify them if texts of close suspects were available.

Published on December 15, 2020, a white paper and a press release (OrphAnalytics, 2020a and b) summarize the results of OrphAnalytics' stylometric analyses: the QAnon corpus forms two clusters in the multivariate analyses. The two groups of messages, Q1 and Q2, appear in chronological order and correspond respectively to the texts published on 4chan, and then on 8chan. In addition, Q2 has been subdivided according to time. QAnon's posts were written over three years: Q1 orange from Oct 28, 2017 to Dec 1, 2017, Q2 red, from Dec 1 2017 to Jun 1, 2018, Q2 magenta from Jun 1, 2018 to Oct 8 2018, Q2 blue from Oct 8, 2018 to Jan 8, 2019, and Q2 cyan from Feb 11, 2019 to Nov 13, 2020.

The two clusters which appear using unsupervised machine learning by multivariate analysis of trigrams are clearly distinct. These stylometric results most likely reveal that at least two authors (at least one author for each of the two style groups) wrote the entire QAnon texts divided into two style clusters, thus shattering the myth of QAnon sympathizers convinced that only one author wrote the entire corpus.

At the time of the press release, few reference texts were available from the six candidates closest to QAnon's management, making any author attribution difficult. Thus, the results of the OrphAnalytics white paper (OrphAnalytics 2020a) do not reveal whether one or two of the six QAnon message managers were the authors of one of the two Q-drops clusters. Therefore, the authors of the two clusters of styles observed in the QAnon corpus could not be identified in that white paper.

However, the finding of two-style groups was validated in an interview in the 6-episode documentary *Q: Into the Storm* (Hoback, 2021) of HBO directed by Cullen Hoback with Paul Furber, an early manager of the corpus publication who observes a change in style associated with the shift of the publication of Q-drops from the 4chan site to the 8chan site (3rd episode of the documentary at 41 min).

Furthermore, in the sixth and final episode of the documentary, Ron Watkins indirectly admits to being Q in an interview, and then publicly recants. This partial confession of Ron Watkins confirms him as the most likely writer of the majority of Q-drops. Before this indirect confession, Ron Watkins was already suspected of writing the Q-drops of Q2. As IT manager of 8chan, he had access to the data of the users who could only publish messages authenticated by tripcode: by his functions, he must have either known the identity of Q, or he was Q himself.

Interestingly, Ron Watkins' resignation as 8chan's IT manager coincides with the cessation of QAnon's publication – the 4953rd and last drop of Q was published on November 13, 2020 – and thus suggests that Ron Watkins was deeply involved in writing QAnon.

With the availability of the texts of the six protagonists involved in managing the publication of Q-drops, it is now possible to compare the styles of the six protagonists with the two styles present in the QAnon corpus in order to determine which of the six write in a style closest to Q1 or Q2.

Available as a service or as the software PATOA standing for [P]rogram of [A]nalysis of [T]exts by [O]rph[A]nalytics or the question in French "Pas toi?" standing for "Not you?", the stylometric techniques used by OrphAnalytics in these analyses are used in forensic investigations and expert reports. More than fifteen investigative or forensic stylometric analyses - more precisely eight analysis reports since the beginning of 2021 - have been produced by OrphAnalytics for academic institutions, investigative firms and private or public courts.

Currently, OrphAnalytics' most publicized forensic expertise concerns a very high-profile untried cold case of France: the 1984 murder of 4-year-old Grégory Villemin known as *l’Affaire Grégory*. Anonymous letters marked this case: the first letters threatened the family, then a message claimed the murder, and finally other mailings taunted the investigation services.

The team of OrphAnalytics was able to reveal to the investigators, with more than 90% certainty, who is the most likely writer of the crime letter whose analyzed transcription contains less than 130 chr. This stylometric strategy, which is able to attribute a likely author on very short text transcriptions – down to a size of at least 50 chr – also makes it possible to establish whether one or more writers participated in the drafting of an anonymous letter with an unparalleled resolution of about 50 chr. This analysis of short messages by what we defined as *stylometric sequencing* has made it possible to establish with a high probability that a single writer conceived the essential criminal messages in the murder of Grégory Villemin.

In addition, OrphAnalytics' stylometric approach was recently validated by a blind analysis of short electronic texts of a criminal case solved by an international police agency. Without any prior knowledge of the context, the PATOA algorithms classified criminal emails of a few hundred characters according to the authors with a precision rate exceeding 90%.

In general, the OrphAnalytics stylometry approach was inspired and adapted from genomic sequence analysis tools. This stylometric sequencing approach has been validated by the analysis of different literary corpora. The analysis of transcripts of the *Millennium* novels written by two authors validated the possibility of measuring two different styles documented in the same novelistic framework (Telesca, 2016). On the other hand, PATOA's analysis of the novels published under the pseudonym of Elena Ferrante, attributing these texts to Domenico Starnone (Genilloud & Roten, 2016), confirmed the results of two physicists from the Sapienza University of Rome, Vittorio Loreto and Andrea Baronchelli, who used text compression to attribute authorship (Delbecq, 2002). Their results were announced in the media (Galella, 2006). These two author attributions by text compression and stylometric sequencing, as well as a third one performed by Tuzzi, & Cortelazzo, were confirmed at the expert meeting on the attribution of Elena Ferrante's novels held in Padua in 2017 (Tuzzi, & Cortelazzo, 2018).

Stylometric analyses were successfully used to assess the authenticity of texts and oral transcriptions of the senate auditions of Brett Kavanaugh and Christine Bradley Ford (OrphAnalytics, 2018).

Material and Methods:

QAnon texts were collected (qresearch.ch) and prepared by RS, as well as the public texts of the authors mentioned in the Q-drops collected by FC, FT, RS, and VR. This corpus was compiled for

stylometric analysis and is available upon request for research purposes only:
contact@orphanalytics.com.

The QAnon texts are those of the 2020 White Paper (Orphanalytics, 2020a): quotations, biblical passages, speeches of US presidents (e.g. Ronald Reagan), links and greetings have been discarded to keep only the parts of the text with a personal syntax. These texts thus correspond to those analyzed in the first stylometric analysis of the QAnon corpus by multivariate analysis and Manhattan distance (Orphanalytics, 2020a).

The temporal divisions selected for the white paper (Orphanalytics, 2020a) have been retained. The abbreviations of the different clusters or parts of clusters follow the color code of the majority of the figures in the first white paper:

Groups and subgroups of style in QAnon:

Q1 orange or Q1O : 79,575 chr (42450 chr 53.3% of chr questions). Q-drops #1 to #237, from 28-10-2017 to 1-12-2017.
Q2 red or Q2R: 21244 chr (5280 chr 24.9% of quest.). Q-drops #238 to #466, from 1-12-2017 to 6-1-2018.
Q2 magenta or Q2M : 168887 chr (43987 chr 26.0% of quest.). Q-drops #468 to #1827, from 6-1-2018 to 10-8-2018.
Q2 blue or Q2B : 227142 chr (76989 chr 33.9% of quest.). Q-drops #1829 to #3570, from 10-8-2018 to 1-8-2019.
Q2 cyan or Q2C : 161821 chr (56707 chr 35.0% of quest.). Q-drops #3574 to #4952, from 2-11-2019 to 13-11-2020.

Reference Texts of Suspects:

Paul Furber: 173,811 chr (8,457 chr 4.9% of quest.): two clusters appear by multivariate analyses according to the white paper (Orphanalytics, 2020a): PaFA and PaFB. The texts of the two Furber styles are used separately for comparisons with QAnon and the other suspects:

PaFA: 50,541 chr (6,368 chr 12.6% of quest.): *Transcription of the Interview Samizdat 36,129 chr, GAB 1,729 chr, Obituary F. Heydenrych 1,712 chr, 8chan 2,539 chr, Discord 953 chr, Questions 7,442 chr.*

PaFB : 123,270 chr (2089 chr 1.7% of quest.): *Big Dick Anon 123,270 chr.*

Tracy Diaz or TraD: 103,033 chr (6,377 chr 6.2% of quest.): *Steemit 41,553 chr, Tweeter 33,952 chr, GAB 27,528 chr.*

Coleman Rogers or CoIR: 11,989 chr (238 chr 2.0% of quest.): *Discord Leaks 9,296 chr, Discord Whispers 514 chr, Patriot Soapbox 1,174 chr, Reddit 1,005 chr.*

Cristina Urso CriU, 108,652 chr (5,658 chr 5.2% of quest.): *text of 14,912 chr, Gab 32,118 chr, Soapbox 61,622 chr.*

Jim Watkins or JimW, 72,944 chr (3,753 chr 5.1% of quest.): *5ch 552chr, Parler 54,895 chr, tweets 16,898 chr, Gab 599 chr.*

Ron Watkins or RonW, 360,511 chr (39,599 chr 11.0% int.).

Text preparation: the corpus was transcribed into text sequences of 27 kinds of characters: 26 letters of the alphabet and the space character to mark the boundaries between each word. Texts of a group are concatenated and trimmed by targeting the size of a 5'000-chr or 20'000-chr sequence with the tolerance needed to obtain an integer number of text fragments from the QAnon corpus.

Simplicity Measure of a Text:

The complexity measure of 20'000-chr sequences of texts of different styles is established from two parameters that are represented in a bidimensional graph. The horizontal axis represents the proportion of complex words in each sequence, i.e. any kind of continuous unit/island of more than 5 letters, limited by spaces. The vertical axis represents the average length of a sentence.

Distance to Barycenters in Trigram Frequency Space

For each sequence of a target size of 5'000-chr, a trigram frequency count is performed in order to represent the sequence by a point in a space of 27^3 or 19683 dimensions, each dimension representing the frequencies of a trigram type. A similarity measure is established for each point representing a 5'000-chr sequence of QAnon and the barycenter of each of the seven style groups.

The approach used in this report differs from that used in the white paper (OrphAnalytics, 2020a): in the absence of dimensionality reduction by multivariate analysis (because we have different texts genres); and in the use of the squared Euclidean distance instead of the Manhattan distance or cos theta used for similarity measurement. Slightly better than the latter, the squared Euclidean distance can be used to strengthen the effect of longer distances in cluster analysis.

It should be noted that multivariate analyses with the squared Euclidean distance on the QAnon corpus give the two clusters, as described in the first white paper (OrphAnalytics, 2020a), which used the Manhattan distance.

Precision and Recall

The effectiveness of the barycenter distance approach is assessed (Figures 2 to 5) by confusion matrix (actual class vs. predicted class).

Results and Discussion

The first white paper showed the presence of two styles in the QAnon corpus (OrphAnalytics, 2020a). When compared in a multivariate analysis, the texts separate into two clusters according to their chronological order. In an interview (Hoback, 2021, 3rd episode at 41st minute) Paul Furber acknowledges a change in style, corresponding to the transfer period of the QAnon corpus, precisely at Q-drop #466, corresponding to the end of Q2R, i.e. the first brief period of Q2.

Multivariate analyses capable of performing authorship attribution by superimposing clusters are not able to establish unequivocally who in the group of suspects writes with a style highly similar, practically identical, to the two Socratic styles Q1 and Q2 found in QAnon.

Syntax analysis: proportion of interrogative sentences

A syntax analysis showed a notable difference in styles between Q1 and Q2: Q1's style being marked by 53.3% of the sequences in the interrogative form, Q2's style using less interrogative forms (31.6%) than that of Q1: respectively for Q2R (24.9%), Q2M (26.0%), Q2B (33.9%), and Q2C (35.0%). The proportion of interrogative sequences characterizing Q2 (31.6%) is significantly higher than in a personal style (e.g. blog, tweets...). The purpose of a Socratic type of writing for Q1 and Q2 is to make the reader permeable to new ideas brought indirectly by questioning.

Among the suspects, Ron Watkins' texts have the highest rate of interrogative sequences of the reference texts of the group of six suspects, i.e. 11.0%, a very high rate of questions in the texts of the six suspects, reflecting the tendency of Q2 to be enriched in interrogative sequences: 31.6%. The same reasoning could be held for Q1 carrying 53.3% interrogative forms. However, Paul Furber seems to be a good candidate as author of Q1, even though he has a lower or equal rate of interrogative sequences as Ron Watkins (total 4.9%, with PaFA 12.6% and PaFB 1.7%): for his

questions list shows that he is capable of writing in a similar questioning style to the important one in Q1 (53.3%).

If a preliminary attribution of authorship is to be made from the maximum proportions of interrogative forms, the case of Ron Watkins as the author of Q1 and Q2 seems to be the most robust. Such rough drawing rules would give 66.7% of the draws for Ron Watkins (11.0% questioning) and 33.3% for Paul Furber, as the latter has two style groups with very different questioning rates: the first group, PaFA, is question-abundant (12,6%), whereas the second group, PaFB, is not (1.7%). For Ron Watkins, such a draw by maximum questioning represents a 4-fold gain compared to a random draw (1/6 or 16.7%). However, even with a lower draw proportion by questioning, Paul Furber cannot be totally excluded as a possible author of Q1 and Q2. A draw based on maximum questioning represents a 2-fold gain for Paul Furber over a random draw (16.7%).

Text complexity measurement

A complexity analysis of texts is illustrated in Figure 1 where each sequence is represented by a point with on the x-axis the proportion of complex words, i.e. of more than five letters, vs. on the y-axis the average length of the sentences of the sequence.

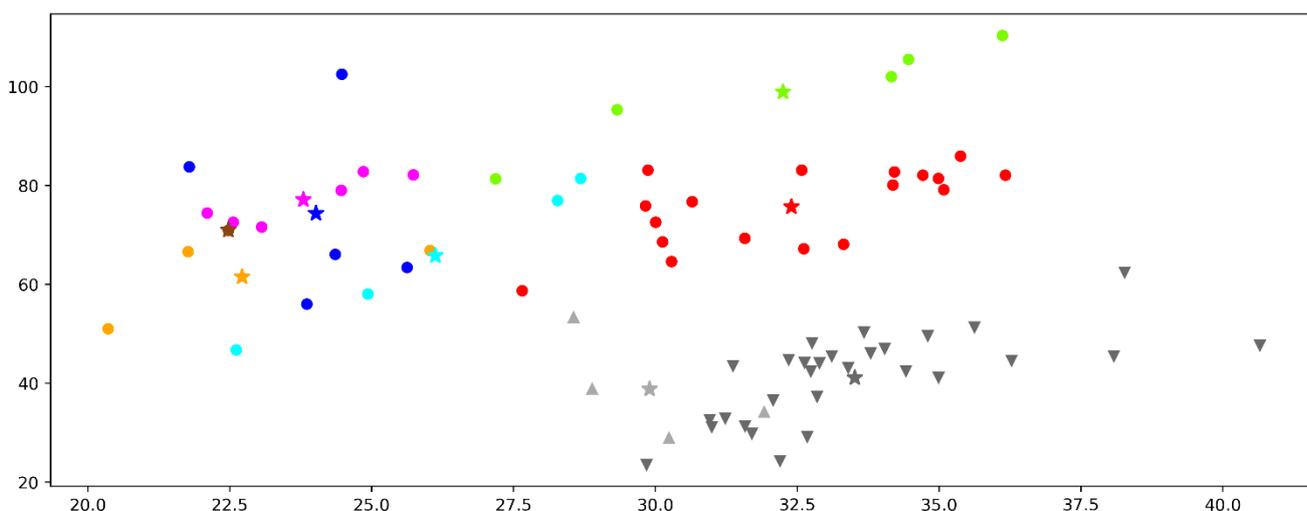


Figure 1: Text complexity of QAnon vs. suspect texts: The text sequences used have a target size of 20'000 chr. Each text sequence is defined by two coordinates. In x, the proportion of complex terms, i.e. more than 5 characters, and in y, the average number of characters per sentence. The texts of Q1 (light grey diamonds) and Q2 (dark grey squares) are compared to the texts of the suspects: Furber (PaFA and PaFB), Diaz, Rogers, Urso, Jim and Ron Watkins (orange, magenta, blue, brown, green, cyan and red circles, respectively). For each color group, a barycenter is calculated and represented by a star.

In Figure 1, few differences between the Q1 and Q2 clusters are visible. These two clusters differ from the 6 suspects' texts by their short sentences and a high proportion of complex words:

- Q1 sentences with 40 characters and 30% of complex words of more than 5 letters,
- Q2 sentences with 40 characters and 34% of complex words of more than 5 letters.

In the suspect group, the analyzed texts are distributed in a continuum from short sentence texts composed of less complex words (average sentence length of 50 chr and 20% of words of more than 5 letters) starting from Paul Furber, Coleman Rogers, Tracy Diaz, and Jim Watkins, to longer sentence texts enriched with more complex words (average sentence length of 95 chr and 36% of words of more than 5 letters) used by Cristina Urso (average sentence length of 95 chr and 36% of words of more than 5 letters) and Ron Watkins (72-chr sentences and 32.5% of words longer than 5 letters).

Only Ron Watkins' texts are closest to those of QAnon in their use of a similar proportion of complex words in rather short sentences. The particular writing of QAnon corresponds to a formatted writing where a significant proportion of parts of sequences using function words have been discarded, in order to enrich the sentences with complex terms (Florian Cafiero, personal communication). To summarize, for the same vocabulary complexity, QAnon texts are written with a simpler syntax than those of the suspects.

QAnon's Socratic writing by questioning requires work on syntax while maintaining semantic richness. Ron Watkins is the author among the six suspects who would have the least difficulty to adapt his writing to that of QAnon by simplifying his syntax while maintaining a similar richness of vocabulary to that of QAnon. The reverse task of keeping the syntax simple while enriching the vocabulary complexity is much harder to achieve (Florian Cafiero, personal communication).

According to this observation, Ron Watkins is thus, among the six suspects, the one who can transform with the least effort his style into that of QAnon's messages by keeping his semantic richness, he would only have to reduce a certain proportion of function words to shorten his sentences. This remark is particularly valid for Q2.

For Q1, if the syntactic complexity remains the same, the vocabulary richness decreases, suggesting a cooperative writing for the beginning of the corpus between people writing with lower vocabulary complexity (Paul Furber, Coleman Rogers, Tracy Diaz or Jim Watkins) and people with richer vocabulary (Cristina Urso or more probably Ron Watkins in view of his syntactic complexity lower than that of Cristina Urso).

Determination of the Likely Writer of QAnon by Certainty Estimates

Since both Watkins appear publicly during the QAnon switch from 4chan to 8chan, an estimate of the probability of assigning one of the six suspects for writing Q2 was performed. To transform a candidate's style to that of Q2, the same vocabulary complexity of Ron Watkins' and Cristina Urso's style can be kept (highly likely case – see above – estimated at 95% certainty), or strongly enriched (unlikely case estimated at 5% certainty). Such a transformation of style also requires either to keep the same syntactic structure (none of the suspects), or to reduce it a little (probable case estimated at 80% certainty for all suspects except Cristina Urso), or to reduce it drastically (unlikely case estimated at 20% certainty for Cristina Urso). According to these rough estimates, the transformation of the Q2 style from Ron Watkins' style is very likely, from Cristina Urso's, less likely, and from the others', unlikely.

The previous calculation on Q2 can also be performed for Q1. However, since none of the candidates write with the vocabulary richness of Q1, it must either be lowered from the style of Ron Watkins and Cristina Urso (one third of the difference between the group of four suspects and the other

two), or significantly increased for the other four suspects, i.e. Paul Furber, Coleman Rogers, Tracy Diaz or Jim Watkins (two thirds of the difference between the group of four suspects and the other two): respectively, both certainties are estimated at 80% and 20%.

A similar probability of syntactic reduction to that of Q2 (above) can be performed. According to these rough estimates, the transformation of the Q2 style from Ron Watkins' style is likely, from Cristina Urso's style, less likely, and from the style of others, unlikely. However, the effort required by all suspects to enrich their vocabulary opens the possibility of a cooperative writing between representative(s) of styles of different complexity. Incorporating this assumption makes quantifying the probability of scenarios complex, and prevents an easy estimate of probable scenarios.

In summary, the complexity of vocabulary and syntax allow us to designate in the group of six suspects either Ron Watkins as a very likely author capable of transforming his style to Q2 style, or Cristina Urso as a less likely author. A similar estimation for Q1 taking into account the additional assumptions could not be performed, but revealed that cooperative writing of Q2 could have occurred.

Style Comparison by Distance to Barycenters of Reference Groups

The style proximity analyses are developed to minimize any possible bias. Texts are cut into sequences of a target size of 5'000 chr. A trigram usage profile is constructed to represent each sequence as a point in a multidimensional space in which each dimension represents the frequencies of a trigram type. Each sequence is evaluated by comparing the distances between the point of the sequence and each one of the seven reference barycenters.

Used in our similarity analyses, the Euclidean quadratic distance is not a distance *stricto sensu*, but a similarity measure optimized for distant object clustering. In the context of the QAnon texts, the resolution obtained with the quadratic Euclidean distances is slightly better than that obtained either with the Manhattan distances, which are easy to calculate, or with the cos theta angular measures between two vectors in the multidimensional space used to compare the similarity between two sequences/barycenters.

In the histograms of Figures 2 to 5, each 5'000-chr sequence is represented by a bar, the color of which corresponds to the group whose barycenter is closest to the sequence. The height of the bar for each sequence corresponds to the shortest of the seven barycenter distances, normalized to the average of the seven barycenter distances established for the sequence shown by the histogram bar.

Figure 2 shows the analysis of the reference texts of suspects and the messages of Q1 and Q2 style groups from QAnon by distance to barycenter. Examination of the 167 reference texts reveals that this approach is able to overwhelmingly associate the sequences of a style group with its style.

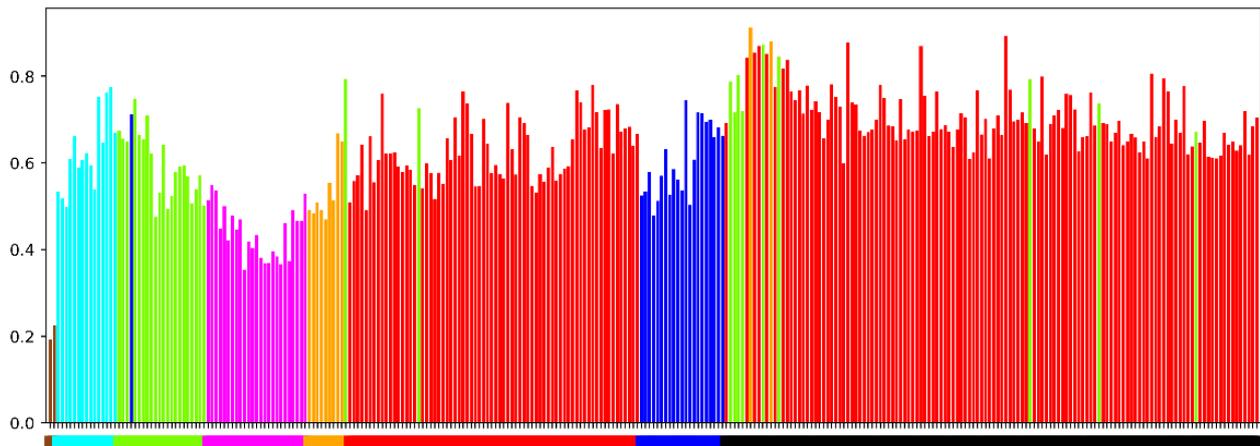


Figure 2: Style comparison of sequences of suspects and QAnon by measures to barycenters calculated with all sequences. The suspects' 5'000 chr sequences appear from left to right: 2 sequences from Coleman Rogers in brown (#1-2), then 15 from Jim Watkins in cyan (#3-17), 22 from Cristina Urso in green (#18-39), 25 from Paul Furber's group B in magenta (#40-64), 10 from Paul Furber's group A in orange (#65-74), 72 from Ron Watkins in red (#75-146), and finally 21 in blue from Tracy Diaz (#147-167). Next, the sequences of Q1 (#168-183) and Q2 (#184-299) follow and are colored according to the closest suspect style. For each of the 167 suspects' sequences, quadratic Euclidean distances are calculated with the seven barycenters of the established reference groups. The shortest distance determines the color of the sequence bar, and its height corresponds to the shortest distance normalized by the average of the distances prevailing between this sequence and the seven barycenters.

The distance-to-barycenter approach measures the links between points in a style group. The shorter the bar, the stronger the link to the style group: e.g. the group of Coleman Rogers in brown and the group B of Paul Furber in magenta. Examination in Figure 2 of all sequences from the seven style groups reveals that only three sequences are not colored according to their group. A precision rate can be computed for the reference groups: 98.2% corresponding to 164 true positives and 3 false positives. This precision is calculated by comparing to the seven barycenters, all calculated with all the sequences of their group.

Unfortunately, this precision is biased for sequences of the same color/style, because this intracluster distance is measured between a sequence and the barycenter of its group calculated with all sequences of the group, including the challenge sequence. The following figures (Figs. 3 to 5) correct for this bias by measuring the intracluster distance between the sequence and this barycenter determined after excluding the challenged sequence.

Taking these caveats into account, the sequences in Q1 and Q2 can already be tentatively assigned to the suspects' style in Figure 2. First, the 16 sequences in Q1 from QAnon are variegated (#168-183), consisting of green, red, and orange sequences while the 116 sequences in Q2 are in red, except for 3 sequences in green. Thus the 16 sequences of 5'000 chr from Q1 appear as a mixture of styles: 8/16 sequences in red are closest to Ron Watkins' style, 6/16 in green to Cristina Urso's and 2/16 in orange to Paul Furber A (more formal style). Next, the 116 sequences of Q2 are overwhelmingly closest to Ron Watkins' style (113 in red, 3 in green).

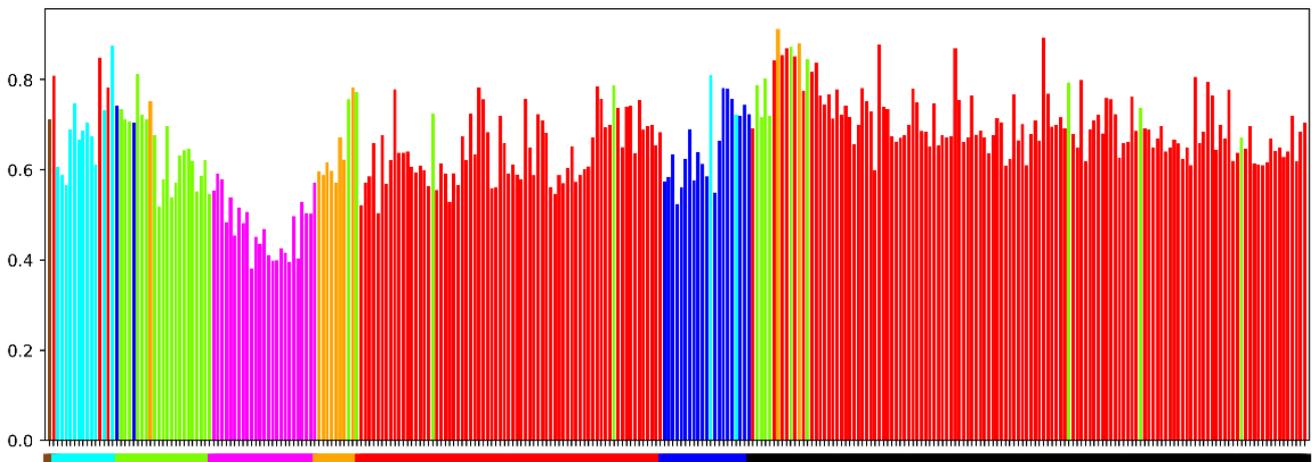


Figure 3: Style comparison of sequences by measure to barycenters calculated without the measured suspect sequence. The conditions of the stylometric analysis of 5'000-chr sequences of the seven suspect styles and Q1 and Q2 are similar to those in Fig. 2. Except for barycenter calculation, the analysis presented in Fig. 3 is identical to the analysis conducted in Fig. 2. This change in intracluster barycenter calculation affects only the height of each bar, and increases the proportion of false positives. The 5'000-chr sequences of the suspects and QAnon appear in the same order and positions as in Fig. 2.

The overall results shown in Figure 3 were established by measures to barycenter, in which the challenged sequence is excluded from barycenter calculation from its group, are similar to those of Fig. 2. This unbiased approach introduces some noise which decreases the computed precision: 92.8%, i.e. 155 true positives and 12 false positives. The style assignment results for Q1 and Q2 in Fig. 3 are essentially identical to those illustrated in Fig. 2.

Since the amount of questions varies between the reference texts and the QAnon texts, which are rich in questions because of their Socratic writing, an experiment similar to that in Figure 3 is conducted on the texts of the six suspects and those of QAnon by separating the non-interrogative sentences (Fig. 4) from the interrogative sentences (Fig. 5).

The overall results shown in Figure 4 established by barycenter measures, in which the challenged sequence is excluded from barycenter calculation from its group, are similar to Fig. 2 and 3. Due to a smaller number of texts than in Fig. 3 (239 vs 299), the computed precision is lower: 90.2%, i.e. 138 true positives and 15 false positives. The style assignment results for Q1 and Q2 in Fig. 4 are identical to those illustrated in Fig. 2 and 3: they show a possible cooperative writing for Q1 (Ron Watkins 5/7; Cristina Urso 2/7), and that the style of Ron Watkins is the closest to Q2 (74/79).

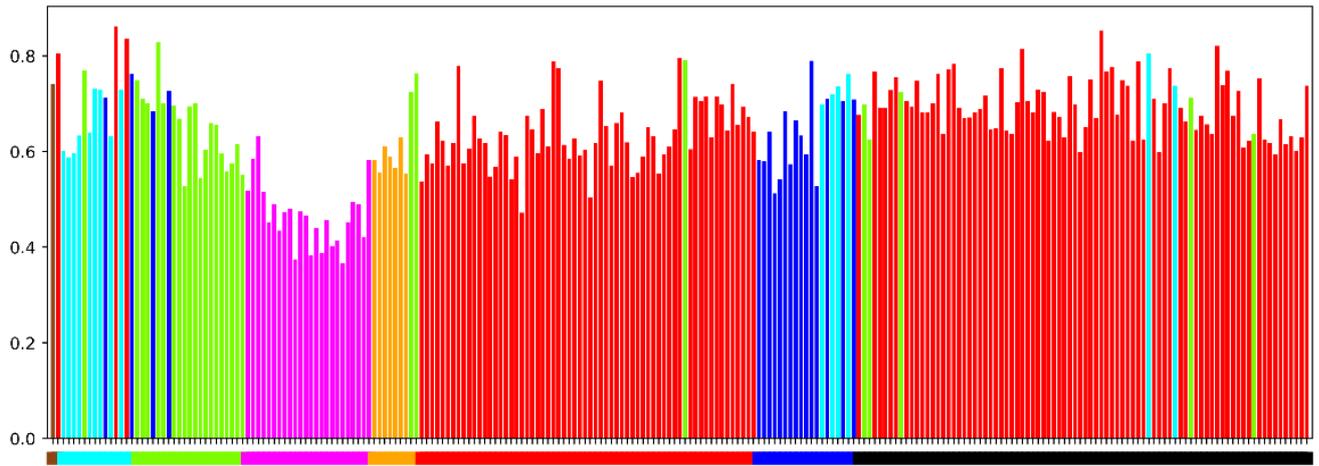


Figure 4: Style comparison of sequences without questions by measure to barycenters calculated without the measured suspect sequence. The conditions of the stylometric analysis of 5'000-chr sequences of the seven suspect styles and Q1 and Q2 (ColR, #1-2; JimW, #3-16; CriU, #17-37; PaFB, #38-61; PaFA, #62-70; RonW, #71-134; TraD, #135-153; Q1, #154-160; Q2, #161-239) are similar to those in Fig. 3. The results are slightly more noisy.

The overall results shown in Figure 5 established by barycenter measures, in which the challenged sequence is excluded from barycenter calculation from its group, are very different from Figures 2-4. Due to a smaller number of texts than in Fig. 3 (59 vs 239 vs 299), only the styles of Paul Furber and Ron Watkins show up in Fig. 5: Ron Watkins as suspect and as QAnon (Q1, 2/8; Q2, 37/37) and Paul Furber as QAnon only (Q1, 6/8). These results confirm the observations that there is cooperative writing in Q1, and that the style of Ron Watkins is the closest to Q2.

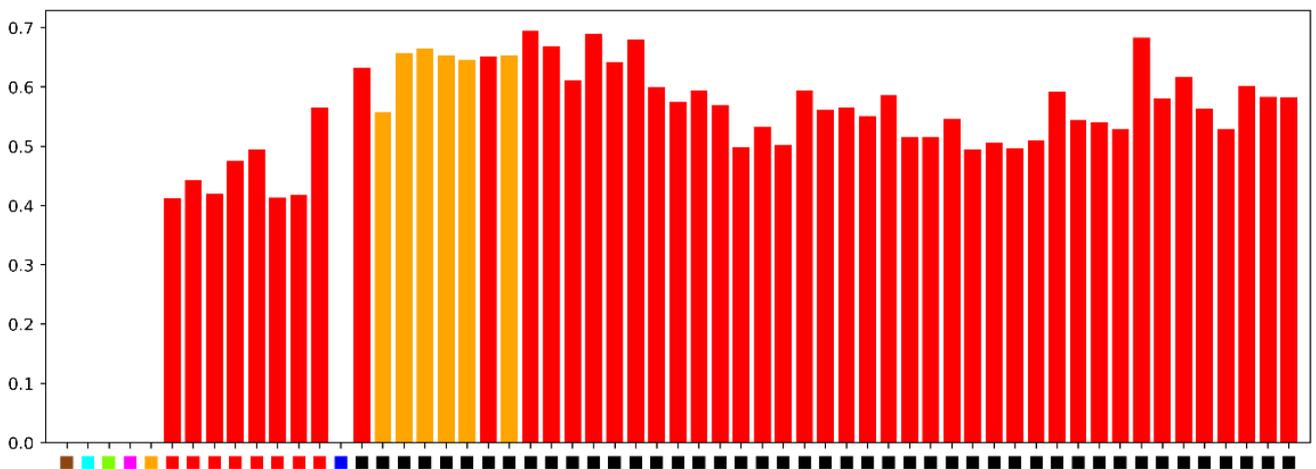


Figure 5: Style comparison of sequences with questions only by measure to barycenters calculated without the measured suspect sequence. The conditions of the stylometric analysis of 5'000 chr sequences of the seven suspect styles and Q1 and Q2 (ColR, no bar; JimW, no bar; CriU, no bar; PaFB, no bar; PaFA, no bar; RonW, #6-13; TraD, no bar; Q1, #15-22; Q2, #23-59) are similar to those in Fig. 3. In order to get a bar, a group of suspects must have at least two 5'000-chr sequences.

Conclusion

First, it should be noted that the ad hoc barycentric approach developed for the analysis of this extremist corpus offers an efficient authentication tool that can be used in large message corpora.

The writing of the QAnon corpus is peculiar: its Socratic style makes questions prominent, sentences are short, and vocabulary is complex. Thus, multidimensional scaled authorship attribution as used in the previous white paper (OrphAnalytics 2020a) is not appropriate for comparing worked Socratic style texts to personal texts of suspects without rewriting. The results of the stylometry experiments presented in this white paper help answer the question of who in the QAnon management group writes with the style closest to the two QAnon's Socratic styles.

In comparisons of writings from suspects and QAnon, the heavy use of questions in Ron Watkins' papers makes his style appear closest to Q2. While Paul Furber can vary his rate of questions considerably, he can spontaneously write a list of questions that make his style most similar to Q1. Both suspects are able to write with an abuse of interrogative forms. With his consistent heavy use of questions, Ron Watkins is the most likely suspect as the author of the bulk of the QAnon corpus, i.e. Q1 and Q2, while Furber's ability to produce a question list favors him as a candidate for Q1, i.e. question-heavy parts of QAnon. These simple observations open up the possibility of cooperative writing for Q1 between Ron Watkins and Paul Furber, and pave the way for a unique production by Ron Watkins for Q2.

Regarding text complexity measured by sentence length and vocabulary complexity, QAnon is written in a peculiar way: short sentences and complex semantics. Since transforming semantics requires more effort than adapting syntax, Ron Watkins is again an ideal suspect for Q2, whereas Q1 appears to be the product of cooperative writing.

The proximity of styles, measured by distance to the barycenter, suggests that Q1 is the product of cooperative writing, with the closest styles being that of Cristina Urso, Ron Watkins, and Paul Furber. The first two styles are close to the non-question parts, with a preference for Ron Watkins, while the last two are involved in the question part, with a clear preference for Paul Furber. Note that in the analyzed cases (complete texts, without or with questions), Ron Watkins always appears in the first analyzed sequence of Q1 (Figures 2-5). A synthesis is possible for Q1: a cooperative writing for sequences without questions close to the style of Ron Watkins' texts and somewhat close to Cristina Urso's, and texts with questions that would be close to Paul Furber's style and somewhat close to Ron Watkins'. For Q2, the answer is clear: Ron Watkins is the most likely author among the six suspects.

Based on these research experiments, QAnon's writing appears at first as a likely collaborative production, and then after a period of time, as the very likely writing of a single author. These results confirm those presented in the first white paper (two clusters).

In summary, it should be noted that the stylometric analysis showing Ron Watkins as the most likely author of the majority of QAnon's Q-drops confirms the impression that Ron Watkins indirectly admitted to being Q before he recanted (Hoback, 2021). Ron Watkins' position as IT manager of 8chan put him in a position to write for Q. For example, Q was silent when 8chan went offline from August 2019 until November 2019, when the site was relaunched and renamed 8kun. Furthermore,

no more Q-drops were published after Ron Watkins resigned as 8kun's IT manager around the last presidential election. This resignation of Ron Watkins from the extremist and pedophile site makes it more politically acceptable. He served in the last days of Donald Trump's administration as an advisor on voter fraud and is currently preparing a candidacy for Congressman of Arizona.

Finding Paul Furber as the author of the first Q-drops seems reasonable. This period is marked by a cooperative style in which the style of Ron Watkins also appears, especially for the very first Q-drop, as well as the style of Cristina Urso, wife of Coleman Rogers and first collaborator of Paul Furber in the management of QAnon. Of the three suspects pinned by stylometry, Paul Furber seems to be the most capable of producing the Socratic style to the Q-drops of Q1. It is conceivable that Ron Watkins and Cristina Urso wrote the first Q-drops, which would have been reworked by Paul Furber. The cooperation between these three authors for Q1 would have given this very peculiar style of QAnon.

Initiated by the attempts at authentication by tripcode, then by the transfer of 4chan to 8chan, the change in style of QAnon is noted by Paul Furber at the time of the transfer. For him, the style deteriorated. Our results, which show the change from a probable cooperative writing of Q1 to the Q2 style most likely used by a single author, supports Paul Furber's analysis.

Finally, it should be noted that of the six suspects, none formally admitted to being Q. If for these suspects being Q means having written all the Q-drops of QAnon, the refusal to admit to being Q is legitimate for those who have only partially participated in the writing of this corpus.

In summary, our stylometric analyses, which have already provided striking evidence in criminal investigations or court litigation, revealed in this study who, in the group involved in running the QAnon publication, most likely wrote the Socratic Q-drops. Overall, the stylometric evidence provided by the QAnon analyses reinforces other evidence brought by other types of investigations.

References

- Delbecq, D. 2002. *La linguistique comprimée par le zip*. Libération 13-3-2002.
https://www.liberation.fr/sciences/2002/03/13/la-linguistique-comprimee-par-le-zip_396800/
- Furber, P. 2021. *Q: Inside The Greatest Intelligence Drop In History*.
<https://paulfurber.net/qinside/qinside.epub>
<https://paulfurber.net/qinside/qinside.pdf>
- Galella, L. 2006. *Ferrante è Starnone: parola di computer*. Unità, p. 26, 23-11-2006
<https://gruppodilettura.files.wordpress.com/2006/11/ferrante.pdf>
<https://www.orphanalytics.com/fr/news/unita-ferrante-e-starnone>
- Genilloud G., Roten C.-A. 2016. *Determination by stylometry of the probable author of the Ferrante corpus: Domenico Starnone*. OrphAnalytics Press Release, 11-10-2016
<http://www.orphanalytics.com/en/news/ferrante-starnone>
- Hoback C. 2021. *Q: Into The Storm*, A six-part documentary series, HBO.
<https://www.hbo.com/q-into-the-storm>

QAnon: Authorship Attribution in a Group of Six Suspects

OrphAnalytics 2018. *Stylometry corroborates Christine Blasey Ford's statements authenticity*. Press Release, 4-10-2018. <http://www.orphanalytics.com/en/news/press-release-october-4-2018>

OrphAnalytics 2020a. *Style analysis by machine learning reveals that two authors likely shared the writing of QAnon's messages at two different periods in time*. White Paper, 15-12-2020. <https://www.orphanalytics.com/en/news/whitepaper202012/OrphAnalyticsQAnon2020.pdf>

OrphAnalytics 2020b. *QAnon is two different people, machine learning analysis shows*. Press Release, 15-12-2020. [http://www.orphanalytics.com/en/news/tag:press release](http://www.orphanalytics.com/en/news/tag:press+release)

Telesca, S. 2016. *Ghostbusters*. Tages Anzeiger 24-2-2016 <https://www.orphanalytics.com/en/news/tages-anzeiger-ghostbusters>

Tuzzi, A., Cortelazzo, M. A. 2018. *Drawing Elena Ferrante's Profile, Workshop Proceedings, Padova, 7 September 2017*. Padova University Press. <http://www.padovauniversitypress.it/publications/9788869381300>

Wang A. & Click S. 2020. *We analyzed every message ever posted by 'Q,' the enigmatic persona that started the QAnon conspiracy theory*. November 3, 2020, Business Insider, <https://www.businessinsider.com/every-QAnon-message-q-drop-analyzed-2020-10>

Zadrozny B. & Collins B. 2018. *How three conspiracy theorists took 'Q' and sparked QAnon*. NBC report., <https://www.nbcnews.com/tech/tech-news/how-three-conspiracy-theorists-took-q-sparked-QAnon-n900531>

Appendix A. Cullen Hoback's Interview of Paul Furber

In early Jan2018, QAnon is starting to pick up steam. And there's a dispute over who controls the Q screen name. Paul Furber, who's running the board, comes out and says:

"This is a fake Q. This is not the Q we know. Q has been taken over by some other unknown actor. That's not the real Q. Ultimately, I blame CodeMonkeyZ (A/N Ron Watkins) because without his help, fake Q would've gone absolutely nowhere. Because every time he posted, I'd have said, 'You're a LARP!' And I would've banned him and I would've deleted the post. But then CodeMonkeyZ said, 'No, this is the real Q!' which was absolute BS. I knew it wasn't true." (Hoback, 2021 – episode 3/6, 41st min)

Hoback comments:

"The person who had the final say on determining whether or not Q had been hijacked was none other than CodeMonkeyZ. However, there was one other thing that did seem to change: Q's writing style. Q went from writing complex, evolving, Socratic questions to posting shout-outs, sharing obviously doctored photos, and using an increasingly absurd amount of exclamation marks. In fact, an AI analysis by Swiss company OrphAnalytics even showed that there was an obvious style shift between Dec. 2017 and Jan. 2018."

Paul Furber says:

"The style changed completely. I thought, no, something wrong here. Just my radar went off."
(Hoback, 2021 – episode 3/6, 42nd min)

Appendix B. Cullen Hoback's interview of Ron Watkins

Cullen Hoback interviewed Ron Watkins and his father Paul over several years for his 6-episodes documentary *Q: Into the Storm*. Ron Watkins tells Hoback:

"It was basically three years of intelligence training, teaching normies how to do intelligence work. It was basically what I was doing anonymously before, but never as Q."

Watkins then smiles and corrects himself, saying:

"Never as Q. I promise. Because I am not Q, and I never was". (Hoback, 2021 – episode 6/6, 49th min)

Hoback comments:

"See that smile. Ron had slipped up. He knew it and I knew it. And after three tireless years of cat and mouse, well..."

Watkins and Hoback both cackle. But Watkins recovers himself and insists:

"No. Never as Q, I promise... because I am not Q... I never was."