

# Concerns Around Opposition to the Green Pass in Italy: Social Listening Analysis by Using a Mixed Methods Approach

Exoteric notes, poorly written but full of passion 😊

## The Friday rule and its implications

The Friday game - unusual methods or topics, the coolness of exploring.

Concept: keep an open mind, see what's going on around, avoid fossilizing, see if there are different ways to do the same things - and don't get stupid because you get bored. I get bored easily.

Thank goodness Nikola is okay with this, barring periods of urgency, but whatever, that's life.

## How to forge a green pass and incidental findings

July 2021: lots of noise about green pass issue. DL 105 July 23, 2021 just passed, need green pass for a number of things (gyms, pools, movie theaters, ...).

Those who don't have a green pass (because they're not vaccinated - it's July 2021) are starting to wonder how to get one. But how do you get a fake green pass?

Internet rule 2: If it exists, people talk about it on reddit or telegram.

I go around trying to figure out what people know and how they try to get a green pass. Not very useful to be honest, since I'm not motivated enough to throw a few hundred quid to some random scammer just to see what happens.

In the meantime, I study the algorithms (the specs are open source) and actually create some fake green passes - unsigned of course, as I don't have the credentials to do so. Like this one, which if you try to use it seriously always tells the truth.



In the process, I come across an interesting digital community - scared and bewildered people trying to figure out the boundaries of something way too complicated for them. And not all of these people are old farts, quite the contrary.

It looks like it's not an intellect deficit problem. But then what is the problem?

This is a gluttony to me because 1. There is empirical data to study, and 2. These data can shed light on issues of ethical relevance, 3. Maybe we can figure out how to fix it, and 4. It's gonna be a ludicrous amount of fun.

The right 'partner in crime' - choose your co-authors, plan and execute a 6-handed job. When you start a small study with no specific funds to carry on in your spare time you need to have partners who are up to the task - fast and responsive and motivated. Thanks Federico and Nikola. Together we make a cerberus: I can collect the data, do the parsing, write the code for the quantitative analysis, do the qualitative analysis. Federico is a dragon at seeing the storyline - because in the end a scientific article is a story told well with the support of data and theory. Nikola often sees things we don't because we have our noses too close to the data - distant reading, in NLP.

## A new topic: quantity and content of no green pass chats

We want to understand this clearly: what are these people saying? How and how much? Who are they? Why do they want a fake GP? Are they vaccinated? Are they antivaccinationists? What are they afraid of? What are their plans?

A lot of questions that are still poorly organized. Let's start organizing them based on the data we have available.

## Data collection: legal and ethical implications

Legal digression: but can we use this data? The GDPR answers: according to article 6.1 we can process personal data without the consent of the data subjects when such activity is in the public interest.

Research is a public interest activity, but because it's a very broad definition we also want to weigh the pros and cons, the benefits and risks, to decide whether or not to do it.

Pros: if we understand what motivates these people, and if as per the hypothesis their motivation is on the ignorance/understanding/fear spectrum, then we can figure out how to talk to them more effectively. Not just listening to them, but trying to understand them to learn to speak their language

Cons: In a dystopian world, I could use chat data to identify them, download their photos, their gps pins, track them down, sneak up on them and in the dark chain them up. Not cool.

Countermeasure: We must ensure total anonymization of data and reduction of the possibility of re-identification

Caveat: re-identification is always just a matter of time, response, and motivation.

## Why Telegram? Shapiro's 'terrorist dilemma'

Why Telegram? Telegram ticks all the boxes. Slightly fringe movements need two contradictory things. Outreach and privacy. You want to reach a lot of people to recruit them to your cause, but you want to keep your identity private. Telegram groups are a hybrid of a social media and messaging system and work very well for this.

Caveat: groups are public - you just need a link to get into them. The content is unencrypted and it's incredibly easy to download an entire group.

## Measure twice and cut once

I'm lazy and therefore hate to waste time. What do we want to distil from this data?

Understand and describe the concerns of no GPs in Italy, the main topics of discussion and their characterization.

The question defines the method and the dataset! A good question matters more than a good answer - the good answer follows by necessity, if you have the question and method and dataset in order.

What do we need? 1. A system to quickly (=quantitatively) analyze discussions that happen across thousands of messages (172926 paragraphs from 53415 people, too much stuff to read the old way - again, I'm lazy). 2. A way to quickly identify which are the most important messages to read the old-fashioned way (NLP is powerful but not omnipotent and in fact machines don't understand things – for now). 3. A primary dataset of no GP chats. 4. A control dataset to make sure we're not getting fireflies for lanterns.

## Hop hop little pony!

We identify interesting chats for the primary dataset. Three university no green pass groups, northern, central, and southern Italy (university students tend to argue their ideas, we noted) plus a generic no green pass group of national scope.

We capture all data and anonymize it. We remove all usernames, images, geodata, names, surnames, toponyms. A bit complex, since in Italian 'Ora' (now) is a toponym and 'Aria' (air) is a name, so the phrase 'ora d'aria' would become [place] d'[name]. But you can do magic with Python.

I write the code for the analysis - available on Zenodo if anyone wants to play with it:

<https://zenodo.org/record/5534045#.YaiVltDMKUm>. The sw does a number of cool things that we didn't put in the paper, e.g: plot the trend of the chat population (how many enter, how many leave), count how many are active (= write) and how many are passive (= just read), plot the amount of messages per day (interesting, if you see a spike up it means that something happened that day, or that something will happen the next day). Again, when you're writing a paper the thing that matters is storyline and consistency, when you have your nose too far in the data you can lose the big picture and it's critical to confront someone who can get you to step back.

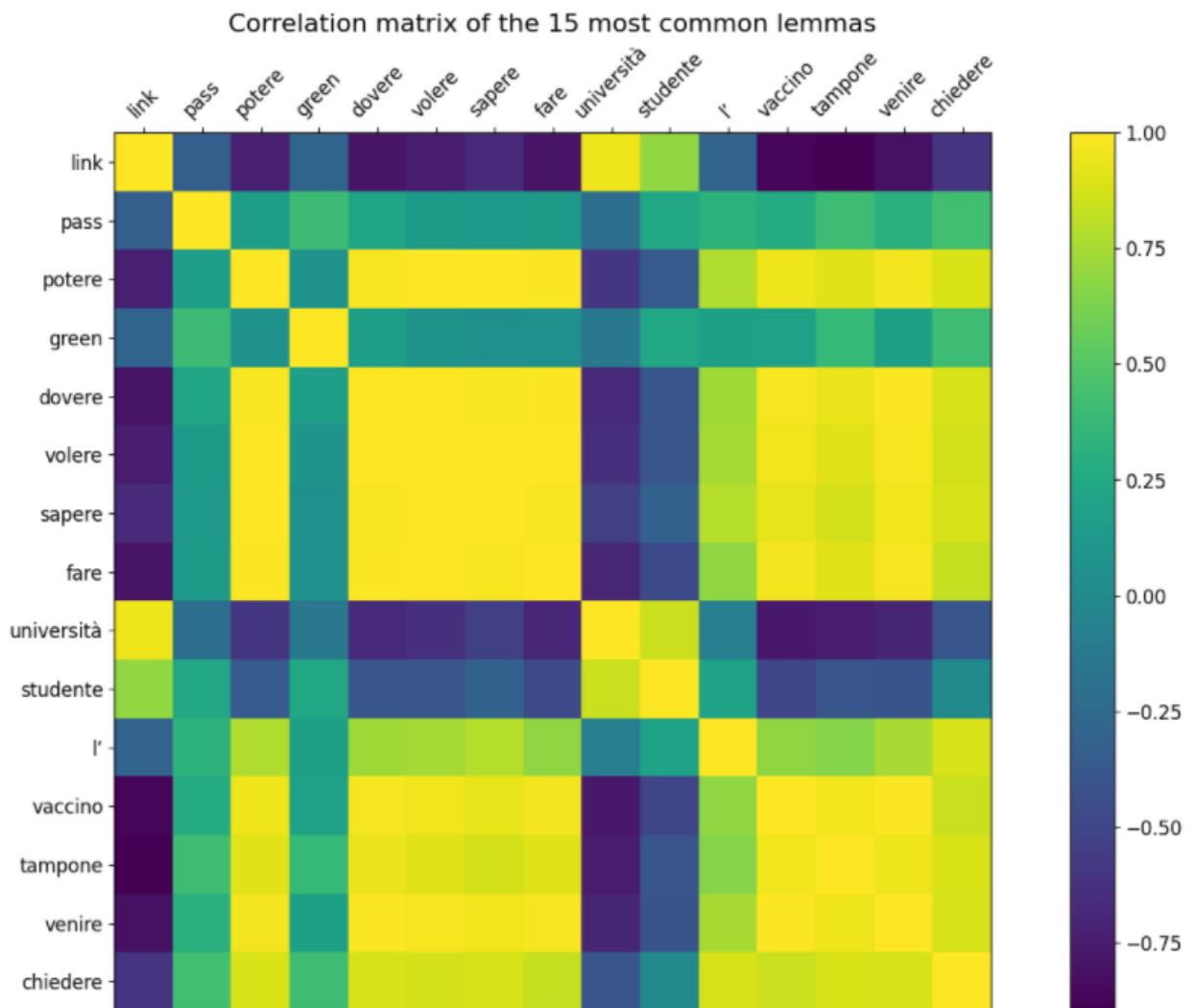
<i>Category</i>	<i>Group description</i>	<i>n of users</i>	<i>n of messages</i>
<i>no green pass</i>	university, north	1770	7356
	university, center	5168	10464
	university, south	479	1879
	generic	12295	33707
	<i>Total</i>	<i>19712</i>	<i>53406</i>
<i>control</i>	parrots	296	48494
	videogames	750	43322
	generic	294	10588
	generic	210	1453
	generic	218	21611

This is what we worked with

## Quantitative Analysis

### Autocoding, or how to make machines do the boring work

The first thing that really matters is autocoding. We wrote a set of rules to identify important concepts to answer our question, e.g. 'c.?o.?v.?i.?d'. All in regex because people aren't dumb and know they're being watched, so they often use codes, like 'the nazipass'. But with a little distributional semantics and regex magic, anything can be done. Distributional semantic analysis reduces words to n-dimensional vectors and calculates the similarity of the vectors, so even if one uses 'pastry' instead of 'green pass' you'll still catch it - because the word is used in the same contexts. There's no code you can't break, it's always a matter of time and motivation.



This is how distributional semantics works, but with fancy colors that are visible also for colorblind people. Green (or yellow) indicates a strong correlation, blue (or purple) a low one.

With autocoding, we have labeled messages, identifying the topics they are talking about and assigning a weight to each label. This is used to see that you're in the right chat and that your rules are working (that's what the control dataset is for).

## NLP: lemma extraction and frequency analysis

A lemma is the root of a word. 'Be' is the lemma of 'am' and 'was'. It is a complexity conflation technique that allows for not sacrificing too much detail.

Lemmatizing a text allows you to count the lemmas and see which ones are most frequent - so what topics people are talking about.

The lemma analysis can obviously be layered, if you have already dissected the corpus into subcorpora (=autocoding). Which words are most frequent in messages that talk about freedom? And which ones in messages that talk about vaccines?

## Sentiment analysis: the temperature of the conversation

The sentiment of a message is a pair of numbers between 0 and 1 whose sum is always 1. It represents the probability of the extender being pissed off. This is done using a neural network and a tensor space, complicated and boring, let's take it as a black box that makes this magic on our behalf - then who wants to read the code is strongly encouraged to do so.

We calculate the sentiment of each message. This allows us to further layer the analysis and ask new questions, e.g: is the sentiment of messages that talk about freedom higher or lower than those that talk about vaccines? Again: with your nose in the data it's easy to get lost, a good philosopher must always remember to step back to see the whole of things and not just the detail.

## Analysis of lemmas and rules: what does it tell us?

The lemmas 'green' and 'pass' are extremely frequent in the no gp dataset, but almost absent in the control dataset. We are not shooting into the wind.

Rules to identify 'freedom', 'vaccine', 'green pass' and 'covid19' fired much more frequently in the no gp dataset. Confirmation, the system works.

The no gp dataset contains many lemmas in the spectrum of acting (can, must, want, pretend, ...)

The no gp dataset contains many lemmas in the law spectrum (law, article, ...)

The 'vaccines' rule is the one that fired most frequently in the no gp dataset. More than the 'green pass' rule.

What can we conclude?

among critics of the green pass, even when the discussion revolves around legal aspects related to personal freedom, skepticism about vaccines probably remains the predominant reason for opposing the green pass - which is interesting, if we keep in mind that they are 'no green pass' and not 'no vax'.

Bonus: In messages tagged as 'green pass', lemmas in the 'vaccine' spectrum are underrepresented. In messages tagged as 'vaccine', lemmas in the 'green pass' spectrum are very frequent.

What can we conclude?

The green pass discussion occurs when vaccines are discussed, but not vice versa. This might suggest that critics of the green pass tend to share anti-vaccine views, but do not want their arguments against the green pass to revolve around their anti-vaccine views. Rather, they prefer to support their position by discussing limitations on personal freedom and advancing legal considerations.

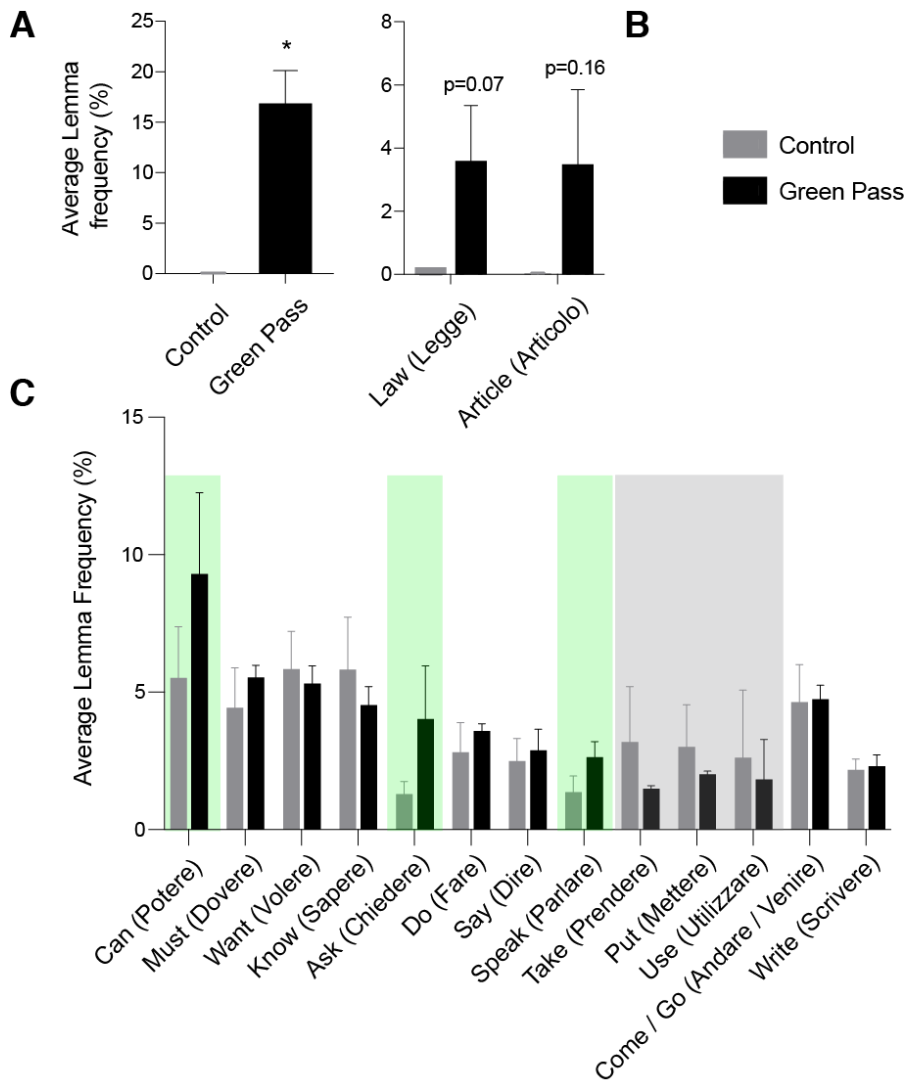


Figure 1 (A) Average lemma frequency (in percentage) in control versus green pass opposition chats. Average lemma frequency (in percentage) for green pass in control (grey bar) versus green pass opposition chats (black bar). (B) Average lemma frequency (in percentage) for legal terms in control chats (grey bars) when compared with green pass opposition chats (black bars), extracted from the 20 most frequently used words in the green pass opposition chats. (C) Average lemma frequency (in percentage) for action terms in control (grey bars) versus green pass opposition chats (black bars), extracted from the 20 most frequently used words in both control and green pass opposition chats. The green background highlights the most relevant action terms that are overrepresented in the green pass opposition chats, whereas the grey background highlights the most relevant action terms that are overrepresented in control chats. \* $P < .05$ ,  $t$  test. Error bars represent standard error of the mean.

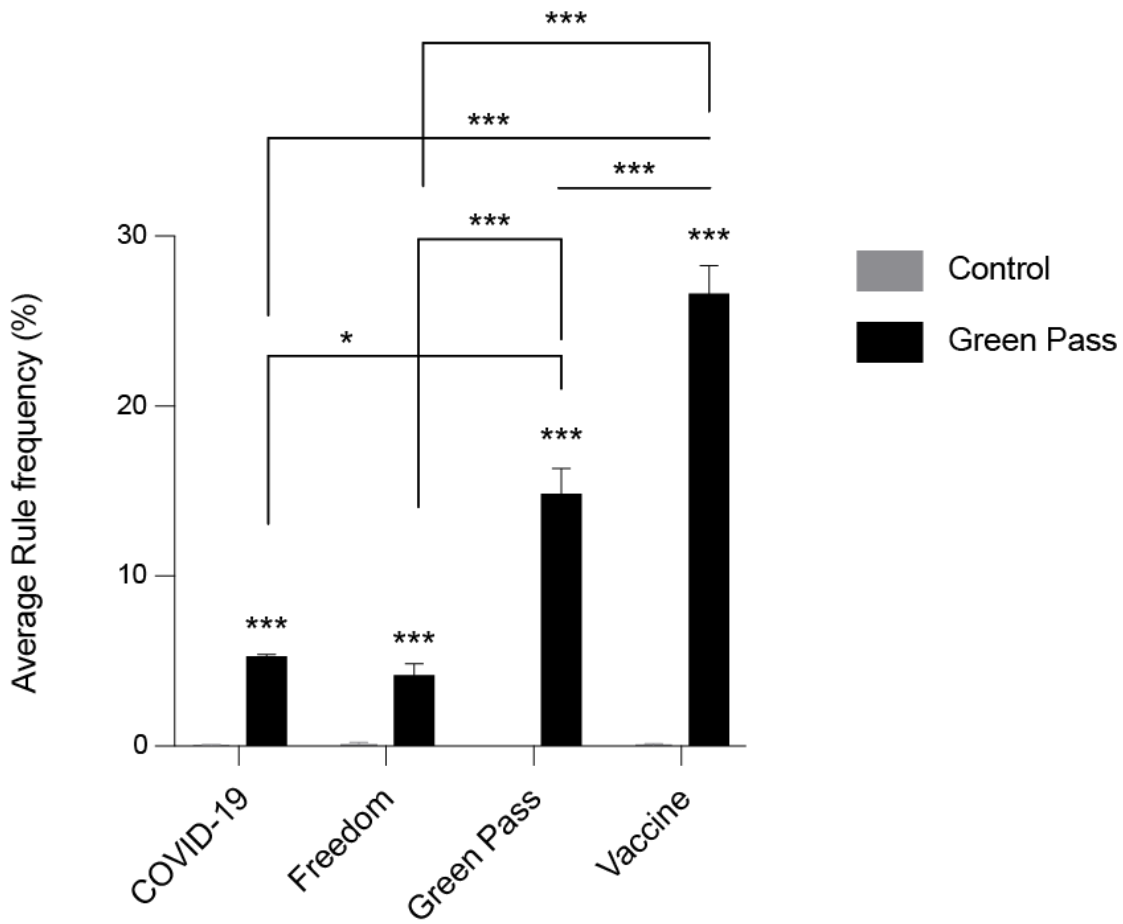


Figure 2. Average rule frequency (in percentage) in control versus green pass opposition chats. Average rule frequency for terms grouped under the rules "COVID-19," "freedom," "green pass," and "vaccine" in control (grey bars) versus green pass opposition chats (black bars). \* $P < .05$ ; \*\*\* $P < .001$ , t test. Error bars represent standard error of the mean.

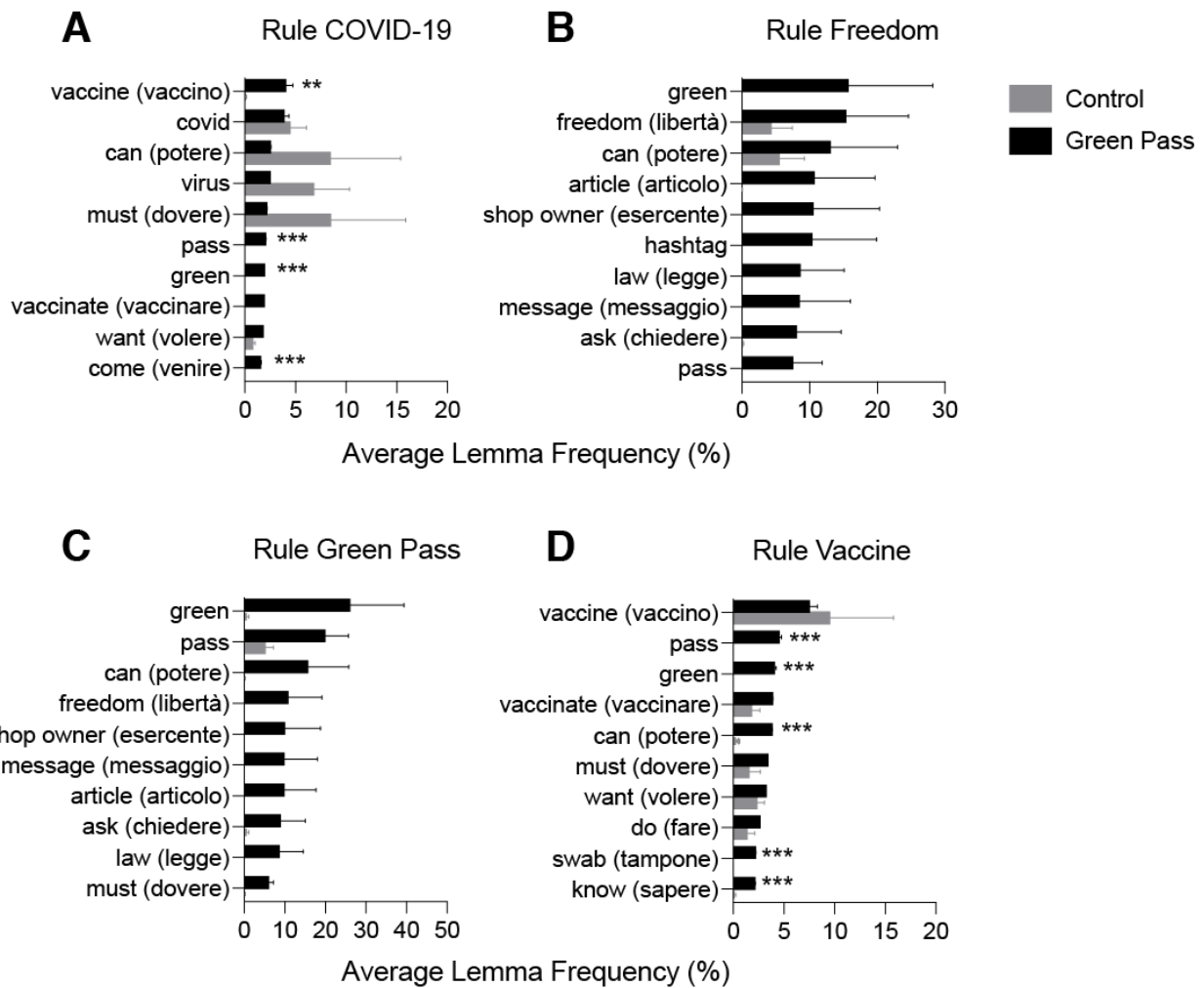


Figure 3. Average lemma frequency (in percentage) per rule in control versus green pass opposition chats. Average lemma frequency in percentage for the 10 most frequent lemmas in green pass opposition chats (black bars), compared with their relative frequency in control chats (grey bars), extracted from messages scoring a positive value for at least one of the rules.

### Sentiment analysis: what do people get pissed off about?

Average sentiment is more negative in the no gp dataset.

The average sentiment in posts tagged as 'freedom', 'vaccine', 'green pass', and 'covid19' is very negative (88%; 96.26%; 85%; 90%)

What can we conclude?

The no gp groups get pissed off about everything, but most of all about vaccines - which is interesting, if we keep in mind that they are 'no green pass' and not 'no vax'.



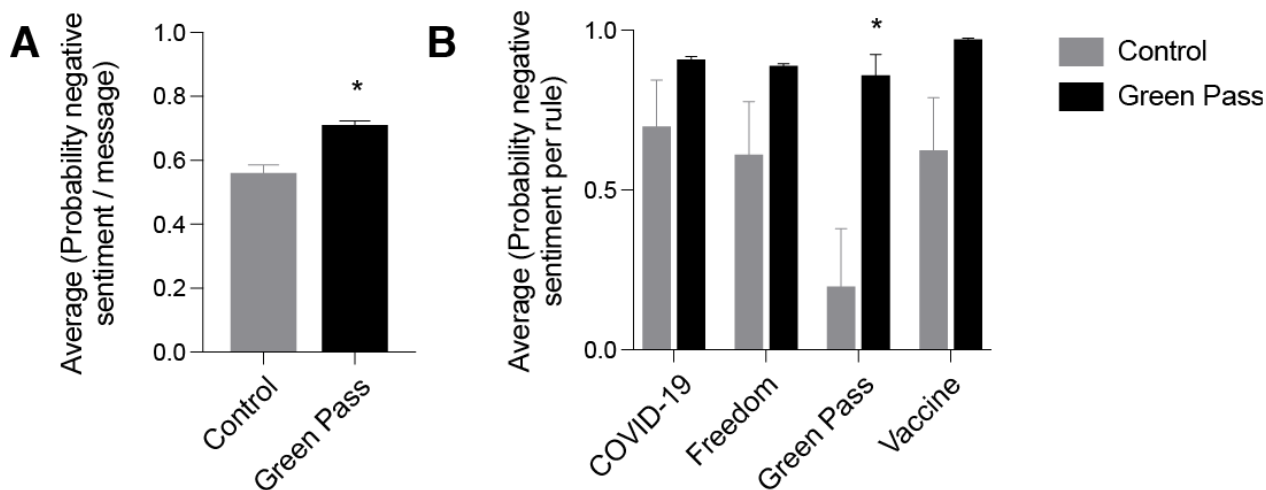


Figure 4. Sentiment analysis in control versus green pass opposition chats. Average probability of negative sentiment in messages published in control (grey bar) versus green pass opposition (black bar) chats. (A) Average probability of negative sentiment per rule in control (grey bars) versus green pass opposition chats (black bars). The following rules are taken into consideration: "COVID-19," "Freedom," "Green Pass," and "Vaccine" (B) 0 indicates the maximum likelihood for an average message to display positive sentiment, whereas 1 indicates the maximum likelihood for an average message to display negative sentiment. \* $P < .05$ , t test. Error bars represent standard error of the mean.

## Qualitative analysis

Two words two about thematic analysis - you take a text, you read it, you label the passages that are interesting to answer your questions, you take all the text belonging to a specific label, you move from vertical to horizontal reading, telling the 'collective story' or 'collective characterization' of a certain theme. You notice and note down the trends, but also the variations.

Aha! But we have a first skim already done. We can use automated tagging (=regex rules) and focus our analysis. This makes the amount of text to analyze more manageable. Lazy and efficient.

Let's throw our noses into the data.

What can we conclude?

Mods ask to stay on topic (= green pass) but users talk a lot about vaccines. 'Can't talk about gp without talking about vaccines'.

Vaccines are scary - because they have side effects, because they don't work, because they are part of an evil plan.

The green pass is illegal and is also that part of an evil plan

Green pass is unethical according to jurisprudential, consequentialist, or deontological arguments.

You have to recognize the 'real enemy' (=not the covid, but 'the politicians') and act with attention and determination.

'the body is mine and I manage it' - this is the main argument, no one can tell me if I should vaccinate or not

Category	Topic	Subcorpus	Pos
Green pass and vaccines	'Stay on topic'	University, south	742

	'How can we ignore vaccines?'	University, north	6693
	'We cannot ignore vaccines'	University, south	1807
		University, center	14716 – 14718
	Critical beliefs about vaccines	University, center	3572 – 3579
	Fear of side effects	University, north	25293 – 25294
		University, center	15682 – 15688
	'Vaccines do not work'	University, north	2612
		University, north	3385
	'Vaccines are part of a bigger scheme'	University, south	2343
		Generic	72471
<i>Beyond vaccines: green pass, legal aspects and personal freedom</i>	'The green pass is illegal and discriminatory'	University, center	3572
		University, center	7520 – 7522
	'The green pass is part of a bigger scheme'	Generic	2127
	Jurisprudential arguments against the green pass	Generic	3448
		University, center	395 – 397
	Consequentialist arguments against the green pass	University, south	3755
		University, south	1255
	Deontological arguments against the green pass	University, center	14996 – 15012
	Converging no-green-pass arguments	Generic	2127
		University, center	5904 – 5906
<i>Action plan</i>	Who are the real enemies?	University, north	20112
	'Avoid defusing topics'	University, north	20233
		University, north	3607
	Protests	Generic	1007
		Generic	2127
<i>Explaining green pass opposition without involving vaccines</i>	'The body is mine and I decide'	University, north	24367
<i>What is COVID-19?</i>	'COVID-19 is just a flu'	University, center	2199 - 2202
		University, north	2864
	'COVID-19 does not exist'	University, north	1328

		University, north	6509
Expertise	'COVID-19 is a means to other ends'	University, center	8092 – 8095
	'Our experts'	University, center	14640 – 14643
		University, south	1974
		University, center	4198 – 4200
Preferred measures		University, north	742
		University, north	3485
	Masks, distancing, tests, dual teaching	University, south	1467
		University, north	25297
		University, center	2095 – 2102
Anti – test and anti – mask positions	Against online teaching and tests	University, north	19204
		University, north	25298
	'Testing is dangerous'	University, north	11697 – 11698
Reliance on anecdotal evidence	'Masks do not work'	University, north	742
	'If it's personal it's true'	University, center	13863 – 13866
		University, north	24524

*Our fancy corpus – in a nutshell*

## Discussion

We have understood and described the concerns of no GPs in Italy, the main topics of discussion and their characterization. And now?

Militant antivaccinism is hardly acceptable, so better to pick on the green pass. The green pass has become the fig leaf of antivaccinationists, and is easier to attack without coming across as dumb. What do we do with this information?

### Recommendations for authorities

1. let's not treat them like fools! Let's acknowledge their fears.
2. let's clarify the purpose of the gp - it's to incentivize vaccinations, not to build the NWO.
3. let's start talking about freedom - what we mean by freedom. Rawls' 'greatest equal liberty' works very well to counter solipsistic / individualistic models of freedom.
4. Explain the legal basis for gp: why it's not illegal and how it fits into the current legal framework and up to when (not for how long) it will be needed
5. let's keep informing about vaccines, with a focus on scientific evidence and transparency.

### Ethical Considerations

The system we have in place works very well. But do we really want it to be a standard? Does the end really justify the means? Always and in all cases? Are there alternatives?

Passive social listening over the long term further destroys fringe groups' lack of trust in institutions. What if instead of spying on them, we asked them directly what they thought? Can we do active social listening? Does it work? Does it allow access to hypercritical positions? Probably yes, we're working on it: <https://www.researchprotocols.org/2021/11/e33653>

## Resources for the curious

These notes: <https://drive.switch.ch/index.php/s/WAm3Ky9Agg7FONE>

This paper: <https://www.jmir.org/2022/2/e34385>

The software: <https://zenodo.org/record/5534045#.YaiVltDMKUm>

Some regex magic: <https://drive.switch.ch/index.php/s/wMwIKZqewFgeQxw>

A test run on a toy dataset: <https://drive.switch.ch/index.php/s/FOAIT3z8wRADu2j>

Toy data for DIY enthusiasts: <https://drive.switch.ch/index.php/s/arv4qZbLqoALBlm>

Sample of actual data included in this study (no GP chat):

<https://drive.switch.ch/index.php/s/WBKanexkDXIeYNv>

Another very cool paper about active social listening: <https://www.researchprotocols.org/2021/11/e33653>

Original messages in Italian: <https://drive.switch.ch/index.php/s/FOAIT3z8wRADu2j>

Original messages in English: <https://drive.switch.ch/index.php/s/MYHIS1WQff5KI8y>



*Goodbye and thanks for all the fish!*