Online search trends and word-related emotional response during COVID-19 lockdown in Italy

Mahmoud M.Elsherif

2023-03-28

### Vignette Setup:

### Project/Data Title:

Online search trends and word-related emotional response during COVID-19 lockdown in Italy

Data provided by: Maria Montefinese

### Project/Data Description:

The strong and long lockdown adopted by the Italian government to limit COVID-19 spreading represents the first threat-related mass isolation in history that can be studied in depth by scientists to understand individuals’ emotional response to a pandemic. The perception of a pandemic threat through invasive media communication, such as that related to COVID-19, can induce fear-related emotions (Van Bavel et al., 2020). The dimension theory of emotions (Osgood & Suci, 1955) assumes that emotive space is defined along three dimensions: valence (indicating the way an individual judges a stimulus; from unpleasant to pleasant), arousal (indicating the degree of activation an individual feels towards a stimulus; from calm to excited) and dominance (indicating the degree of control an individual feels over a given stimulus; from out of control to in control). Fear is characterized as a negatively valenced emotion, accompanied by a high level of arousal (Witte, 1992; Witte, 1998) and a low dominance (Stevenson, Mikel & James, 2007). This is generally in line with previous results showing that participants judged stimuli related to the most feared medical conditions as the most negative, the most anxiety-provoking and the least controllable (Warriner, Kuperman & Brysbaert, 2013). Fear is also characterized by extreme levels of emotional avoidance of specific stimuli (Perin et al., 2015) and may be considered a unidirectional precursor to psychopathological responses within the current context (Ahorsu et al., 2020). dealing with fear in a pandemic situation could be easier for some people than others. Indeed, individual differences have been associated with behavioral responses to the pandemic status (Carvalho Pianowski & Gonçalves, 2020).

To mitigate the COVID-19 effects on individuals’ mental health, it is compelling to evaluate their emotional response to this emergency. Internet searches is a direct tool to address this issue. Indeed, it has been reported that COVID-19 affected the content that people explored online (Effenberger et al., 2020), and online media and platforms offer essential channels where people convey their feelings and emotions and seek health-related information (Kalichman et al., 2003; Reeves, 2001). In particular, Google Trends is an available data source of real-time internet search pattern, which has been demonstrated to be a valid indicator of people’s desires and intentions (Payne, Brown-Iannuzzi & Hannay, 2017; Pelham et al., 2018). Thus, the amounts of COVID-19-related internet searches revealed by Google Trends are an indicator of how people feel about concepts related to the COVID-19 pandemic. A shift in online search trends reflects a change in participants’ interests and attitudes towards a specific topic. Based on the topic, the context (i.e., the reasons causing this change), and this mutated interest per se, it is possible to predict people’s behavior and affective response towards the topic in question. In this study, we aim to understand how emotional reaction and online search behavior has changed in response to the COVID-19 lockdown in the Italian population.

### Methods Description:

Data were collected in the period from May 4th to May 17th, 2020, the last day of full lockdown in Italy, from 71 adult native Italian speakers (56 females and 13 males; mean (SD) age = 26.2 (7.9) years; mean (SD) education = 15.3 (3.2) years). There were no other specific eligibility criteria. An online survey was conducted using Google Forms to collect affective ratings during the lockdown caused by the COVID-19 epidemic in Italy. In particular, we asked participants to complete the Positive and Negative Affect Schedule (PANAS, Terraciano, McCrae & Costa, 2003) and Fear of COVID-19 Scale (FCV-19S, Ahorsu et al., 2020) and judged valence, arousal, and dominance (on a 9-point self-assessment manikin, Montefinese et al., 2014) of words either related or unrelated to COVID-19, as identified by Google search trends. The word stimuli consisted in 3 groups of 20 words each. The first group (REL+) consisted in the words showing the largest positive relation between their search trends and the search trend for the COVID-related terms. By contrast, the second group (REL-) consisted in the words showing the largest negative relation between their search trends and the search trend for the COVID-related terms. In other words, the COVID-19 epidemic in Italy, and the consequent increase in interest for the COVID-related terms, was related to a similar increase of interest for the REL+ words and a decrease of interest for the REL- words. The third group (UNREL) consisted in the words for which the search trend was unrelated to the search trend for the COVID-related terms.

### Data Location:

<https://osf.io/we9r4/>

DF <- import("montefinese\_data.csv")  
  
names(DF) <- make.names(names(DF),unique = TRUE)  
  
names(DF)[names(DF) == 'ITEM..ITA.'] <- "item"  
  
DF <- DF %>%  
 arrange(item) %>% #orders the rows of the data by the target\_name column  
 group\_by(item) %>% #group by the target name  
 transform(items = as.numeric(factor(item)))%>% #transform target name into a item  
 select(items, item, everything()  
 ) #select all variables from items and target\_name   
  
head(DF)

## items item ssID Gender Age Education Measure StimType Response  
## 1 2 Agitato 1 F 36 21 PANAS- 2  
## 2 2 Agitato 2 M 40 21 PANAS- 1  
## 3 2 Agitato 3 F 29 21 PANAS- 2  
## 4 2 Agitato 4 M 39 13 PANAS- 1  
## 5 2 Agitato 5 F 27 16 PANAS- 1  
## 6 2 Agitato 6 M 33 18 PANAS- 1

### Date Published:

2021-08-10

### Dataset Citation:

Montefinese M, Ambrosini E, Angrilli A. 2021. Online search trends and word-related emotional response during COVID-19 lockdown in Italy: a cross-sectional online study. PeerJ 9:e11858 <https://doi.org/10.7717/peerj.11858>

### Keywords:

Covid-19; Emotional response; Online search; Lockdown; Coping

### Use License:

CC-By Attribution 4.0 International

### Geographic Description - City/State/Country of Participants:

Italy

### Column Metadata:

metadata <- import("montefinese\_metadata.xlsx")  
  
flextable(metadata) %>% autofit()

| Variable Name | Variable Description | Type (numeric, character, logical, etc.) |
| --- | --- | --- |
| ssID | Participant code | Numeric |
| Gender | Participants’ gender | Character |
| Age | Participants’ age | Numeric |
| Education | Participants’ years of education | Numeric |
| Measure | Questionnaires and ratings (PANAS, COVID-19 fear, valence, arousal, dominance) | Character |
| ITEM (ITA) | Test items and word stimuli | Character |
| Stim Type | Word condition (REL+, REL-, UNREL) | Character |
| Response | Participants’ scores to the questionnaires and ratings | Numeric |

### AIPE Analysis:

In this dataset, there are REL+ and REL- variable. In REL+ condition, words showing largest positive relation between their search trends and the search trend for the COVID-related terms. In REL- condition, words showed the largest negative relation between their search trends and the search trends for the COVID-related terms.

#### Stopping Rule

What the usual standard error for the data that could be considered for our stopping rule using the 50% decile?

# individual SEs for how surprising   
SE\_full <- tapply(DF$Response, DF$StimType, function (x) { sd(x)/sqrt(length(x)) })  
SE\_full

## REL- REL+ UNREL   
## 0.02795693 0.03026960 0.03377077 0.03497602

quantile(SE\_full, probs = .5)

## 50%   
## 0.03202019

Given the differences in conditions, we subset the data to each condition to estimate separately.

### create subset for REL+  
DF\_RELpos <- subset(DF, StimType == "REL+")  
DF\_RELpos\_re <- cast(DF\_RELpos, ssID + item ~ StimType, value = "Response")

## Aggregation requires fun.aggregate: length used as default

### create subset for REL-  
DF\_RELneg <- subset(DF, StimType == "REL-")  
DF\_RELneg\_re <- cast(DF\_RELneg, ssID + item ~ StimType, value = "Response")

## Aggregation requires fun.aggregate: length used as default

### create subset for UNREL  
DF\_UNREL <- subset(DF, StimType == "UNREL")  
DF\_UNREL\_re <- cast(DF\_UNREL, ssID + item ~ StimType, value = "Response")

## Aggregation requires fun.aggregate: length used as default

# individual SEs for REL+ condition   
SE1 <- tapply(DF\_RELpos$Response, DF\_RELpos$item, function (x) { sd(x)/sqrt(length(x)) })  
  
SE1

## burro casa cioccolato computer corona famiglia febbre   
## 0.1041764 0.1579982 0.1514176 0.1216972 0.1139943 0.1429923 0.1366731   
## lavare libertà mondo muffin notizie peste salute   
## 0.1162521 0.1451916 0.1321747 0.1341764 0.1180695 0.1605518 0.1360807   
## salvare sole tempo termometro torta vaiolo   
## 0.1211320 0.1579153 0.1496335 0.1172752 0.1370362 0.1436611

quantile(SE1, probs = .5)

## 50%   
## 0.1363769

# individual SEs for REL- condition  
SE2 <- tapply(DF\_RELneg$Response, DF\_RELneg$item, function (x) { sd(x)/sqrt(length(x)) })  
  
SE2

## autobus costa dormire giacca hotel mangiare matrimonio   
## 0.11266867 0.13426010 0.17863089 0.09332695 0.10308285 0.13960910 0.13240030   
## motore palazzo pantalone piazza profumo ristorante sposa   
## 0.09077204 0.10040452 0.10719557 0.10986544 0.14308026 0.11724859 0.12092330   
## tram tumore uomo vacanza viaggio villaggio   
## 0.11213589 0.19936188 0.09260898 0.15237180 0.13118463 0.10510344

quantile(SE2, probs = .5)

## 50%   
## 0.1149586

# individual SEs for UNREL condition  
SE3 <- tapply(DF\_UNREL$Response, DF\_UNREL$item, function (x) { sd(x)/sqrt(length(x)) })  
  
SE3

## affogare baco cannone cappio corridore dipendente   
## 0.19629740 0.09809148 0.14121073 0.15099820 0.10197245 0.13519309   
## disturbare fetore firmamento funerale gusto ladro   
## 0.13887361 0.14580113 0.15128780 0.16268821 0.11149057 0.16932199   
## malevolenza mestolo nettare oceano offendersi orgasmo   
## 0.14638612 0.10616543 0.11490455 0.16741016 0.14576833 0.14645215   
## perfezione tradire   
## 0.14609605 0.18521797

quantile(SE3, probs = .5)

## 50%   
## 0.1459486

# sequence of sample sizes to try  
samplesize\_values <- seq(25, 300, 5)  
  
# create a blank table for us to save the values in   
sim\_table <- matrix(NA,   
 nrow = length(samplesize\_values),   
 ncol = length(unique(DF\_RELpos$item)))  
# make it a data frame  
sim\_table <- as.data.frame(sim\_table)  
  
# add a place for sample size values   
sim\_table$sample\_size <- NA  
sim\_table$var <- "Response"  
  
# make a second table for the second variable  
sim\_table2 <- matrix(NA,   
 nrow = length(samplesize\_values),   
 ncol = length(unique(DF\_RELpos$item)))  
  
  
# make it a data frame  
sim\_table2 <- as.data.frame(sim\_table2)  
  
# add a place for sample size values   
sim\_table2$sample\_size <- NA  
sim\_table2$var <- "Response"  
  
# loop over sample size  
for (i in 1:length(samplesize\_values)){  
   
 # temp dataframe for age trait that samples and summarizes  
 temp\_RELpos <- DF\_RELpos %>%   
 dplyr::group\_by(item) %>%   
 dplyr::sample\_n(samplesize\_values[i], replace = T) %>%   
 dplyr::summarize(se1 = sd(Response)/sqrt(length(Response)))   
   
 #   
 colnames(sim\_table)[1:length(unique(DF\_RELpos$item))] <- temp\_RELpos$item  
 sim\_table[i, 1:length(unique(DF\_RELpos$item))] <- temp\_RELpos$se1  
 sim\_table[i, "sample\_size"] <- samplesize\_values[i]  
   
 # temp dataframe for that samples and summarizes  
   
 temp\_RELneg <-DF\_RELneg %>%   
 dplyr::group\_by(item) %>%   
 dplyr::sample\_n(samplesize\_values[i], replace = T) %>%   
 dplyr::summarize(se2 = sd(Response)/sqrt(length(Response)))   
  
 #   
 colnames(sim\_table)[1:length(unique(DF\_RELneg$item))] <- temp\_RELneg$item  
 sim\_table2[i, 1:length(unique(DF\_RELneg$item))] <- temp\_RELneg$se2  
 sim\_table2[i, "sample\_size"] <- samplesize\_values[i]  
}

Suggestions for REL+ Condition:

# multiply by correction   
cutoff <- quantile(SE1, probs = .5)  
  
final\_sample <-   
 sim\_table %>%  
 pivot\_longer(cols = -c(sample\_size, var)) %>%   
 dplyr::rename(item = name, se = value) %>%   
 dplyr::group\_by(sample\_size, var) %>%   
 dplyr::summarize(Percent\_Below = sum(se <= cutoff)/length(unique(DF\_RELpos$item))) %>%   
 dplyr::filter(Percent\_Below >= .80) %>%   
 dplyr::arrange(Percent\_Below) %>%   
 mutate(new\_sample = round(39.369 + 0.700\*sample\_size + 0.003\*cutoff - 0.694\*length(unique(DF\_RELpos$items))))

## `summarise()` has grouped output by 'sample\_size'. You can override using the  
## `.groups` argument.

flextable(final\_sample) %>% autofit()

| sample\_size | var | Percent\_Below | new\_sample |
| --- | --- | --- | --- |
| 250 | Response | 0.80 | 200 |
| 255 | Response | 0.80 | 204 |
| 260 | Response | 0.80 | 207 |
| 270 | Response | 0.85 | 214 |
| 280 | Response | 0.85 | 221 |
| 295 | Response | 0.85 | 232 |
| 275 | Response | 0.90 | 218 |
| 290 | Response | 0.90 | 228 |
| 285 | Response | 0.95 | 225 |
| 300 | Response | 1.00 | 235 |

cutoff <- quantile(SE2, probs = .5)  
final\_sample2 <-   
 sim\_table2 %>%  
 pivot\_longer(cols = -c(sample\_size, var)) %>%   
 dplyr::rename(item = name, se = value) %>%   
 dplyr::group\_by(sample\_size, var) %>%   
 dplyr::summarize(Percent\_Below = sum(se <= cutoff)/length(unique(DF\_RELneg$item))) %>%   
 dplyr::filter(Percent\_Below >= .80) %>%   
 dplyr::arrange(Percent\_Below) %>%   
 mutate(new\_sample = round(39.369 + 0.700\*sample\_size + 0.003\*cutoff - 0.694\*length(unique(DF\_RELpos$items))))

## `summarise()` has grouped output by 'sample\_size'. You can override using the  
## `.groups` argument.

flextable(final\_sample2) %>% autofit()

| sample\_size | var | Percent\_Below | new\_sample |
| --- | --- | --- | --- |

#### Minimum Sample Size

Based on these simulations, we can decide our minimum sample size is likely close to 200 or .

#### Maximum Sample Size

In this example, we could set our maximum sample size for 90% power, which would equate to 218 or participants.

#### Final Sample Size

In any estimate for sample size for this study, the dataset has a large variance in ratings. This dataset need to more sample for items in each conditions. Each condition showed slighty different sample size suggestions, which we should account for.