Appendix 1 - Methodological note

As more briefly described in the main paper, data presented in our study are analysed relying on Computer-Assisted Text Analysis (CATA) techniques, employing two software packages: R studio and T-Lab.

Data were retrieved automatically from Reddit and Youtube while it was necessary to manually identify comments URLs due to Facebook current limitations to its graph API. In the first case, we specifically relied on the «redditextractor» and «tuber» libraries (Rivera, 2019; Sood et al., 2020). For what concerns Facebook, comments were obtained by manually downloading the files of each post, which were then processed with R Studio to anonymise and then extrapolate the individual comment (via the «rvest» and «stringr» libraries). We used R Studio also for the 2nd phase, to run the pre-processing and corpora normalizing activities: we cleaned the texts from spam, links, emojis, and words longer than 50 characters, if any. The data thus obtained were collected in a single file in .csv format.

We then moved to the T-Lab package for Computer-Assisted Text Analysis for the 3rd phase (import, assembling and lexicalisation of *corpora*). We relied on the software preset dictionaries, which have been specifically enriched over the years to pre-process the Italian language: stop-words were removed, whereas multi-words (i.e. idiomatic expressions and locutions) were identified and recorded in the vocabulary of each corpus.

In the 4th phase, we applied to the pre-processed *corpus* the Thematic Analysis of Elementary Contexts (TAEC) provided by the software, which considers as an EC every sequence of word tokens interrupted by the «full stop» (carriage return) and whose dimensions are less than 400 characters Lancia (2012); T-Lab (2021).

The main phases of TAEC analysis are presented below:
a - construction of a «context unit x lexical unit» matrix, with presence / absence values;
b – pre-treatment of data by TF-IDF and transformation of each row vector to length 1 (Euclidean norm); the TF-IDF measure allows to evaluate the importance of a term (lexical unit) within a document (context unit);
c – partition of corpora in ECs (tab. A) using the cosine similarity measure (Weller and Romney, 1990) and the unsupervised clustering method «bisecting K-means» (Karypis et al., 2000; Savaresi and Boley, 2004), accordingly to the exploratory nature of the research.

Tab. A – Key metrics of the ECs classification made by T-Labs software (N total ECs in corpus = 16,738).

Cluster N	N of ECs classified	% total classified (% total in the corpus)
1	2,152	16.1 (12.8)
2	3,335	24.9 (19.9)
3	4,523	33.8 (27.1)
4	1,358	10,1 (8,1)
5	2,023	15.1 (12,1)
Total ECs classified	13,391	100,0 (80,0)

For each of the partitions obtained:

d. construction of a contingency table of «lexical units x cluster» (n x k);

e - chi-square test applied to all cross«cluster x lexical units», to have a measure of the lexicon peculiarity and minimize biases based on the researcher subjectivity when picking up text excerpts for the qualitative discussion (Krippendorff, 2004; Baron et al., 2009; Baker, 2006).

f - correspondence analysis of the contingency table of «lexical units x cluster» (Benzécri and Benzécri, 1984; Greenacre, 1984; Lebart et al., 1998).

T-Lab's TAEC is implemented following an algorithmic logic based on a double reiteration of the process detailed below (for further details, see T-Lab, 2021; Lancia, 2012):

1. Calculation of the «Intraclass-Correlation Coefficient» (ICC-rho in table B below), which corresponds to the ratio between intercluster variance and total variance;

2. measurement of the «gap», i.e. difference between the value of the ICC-rho coefficient and that of the immediately preceding partition;

3. crosscheck of the partition results via two indexes:

    a. the Caliński and Harabasz (1974) index, which shows higher scores when clusters are dense and well separated;

    b. the Davies and Bouldin (1979) index, which, with an opposite logic, is based on a ratio of within-cluster and between-cluster distances and thus shows lower values when clusters are dense and well separated.

4. the production of a child partition is stopped when the gap (see point 2 above) measured after the *n-th* attempt starts to decrease;

5. After controlling for the indexes in point 3, the software by default suggests the (n-1) attempt as the best-fitting partition model from a statistical point of view.

It is worth noting that the choice of the optimal solution is also the result of an interpretation of the outputs (Lancia, 2012; Franzosi, 2008), considering qualitative aspects such as, to name a few: the semantic nature of the clusters; the theoretical justification of their spatial layout; possible overlapping among some clusters, if any.

Considering all the above mentioned quali-quantitative factors, the authors unanimously agreed to opt for the 5-partitions model highlighted in tab. B below:

Tab. B – Key metrics of the TAEC made with T-Lab using its unsupervised clustering algorithm (N attempts = 7).

Partitions/attempts	ICC (rho)	Gap	Caliński-Harabasz	Davies-Bouldin
2	.003	.000	-	-
3	.007	.004	45.894	48.619
4	.012	.005	52.898	21.090
5	.018	.006	61.397	10.901
6	.027	.009	74.254	6.009
7	.033	.006	76.613	4.176
Selected solution: 5 clusters Cosine similarity: .039				

Generally speaking, the TAEC is a multivariate method used to reduce the multidimensionality of one or more *corpora* to fit a Cartesian plane. To do so, data from a matrix in the form of «words x EC» are grouped into clusters on a bi-dimensional plane, where the x-axis accounts for the *inertia* of the table – which is the maximum amount of association along the horizontal axis (Weller and Romney, 1990) – and the y-axis «seeks to account for a maximum of the remaining association» (Schonhardt-Bailey, 2008: 403). In other words, this means that occurrences or clusters showing «similar distributions will be represented as points that are close in space, and categories that have very dissimilar distributions will be positioned far apart» (Clausen, 1998: 2; also see Anstead, 2018).

TAEC can be interpreted in three or more dimensions, although this study is limited to a two-factor design analysis. The axis not discussed in the main article are nevertheless reported in tab. C:

Tab. C – Thematic Analysis of Elementary Contexts (TAEC) results.

Factor	Eigenvalues	%	Cum. %
1	.227	37.44	37.44
2	.165	27.16	64,6
3	.119	19.54	84,14
4	.010	15.86	100.00

At the end of the process, the technique provides a mapping of homogeneous semantic clusters (i.e. isotopies) into which one or more corpora are divisible, and which can legitimately be interpreted as a representation of the «general or specific themes» which inform the textual material under scrutiny (Rastier et al., 2002). For our investigations, we selected as a metric the «cosine similarity» (Huang, 2008; Mihalcea et al., 2006) and the «bisecting K-means» as a partition method (Karypis et al., 2000).

Lastly, it is worth noting that the x- and y-axes are not labelled by the software. Their semantic values are abductively determined by the researcher using an appropriate theoretical framework of interpretation and argumentation (for a detailed discussion, see Cardano, 2020; a general framework for the theory of interpretation is discussed in Eco, 1992). The possibility to mix quantitative and qualitative approach in the interpretation of semantic material «is the great strength of correspondence analysis, as it has the power to reveal the underlying structure of the data» (Anstead, 2018: 294).

For a more detailed presentation of the CA family of techniques and a discussion on its practical applications, see Greenacre (2017); Clausen (1998). To have a full account on how T-Lab implements TAEC, see T-Lab (2021); Lancia (2012).

REFERENCES TO APPENDIX 1

Anstead N (2018) The idea of austerity in British politics, 2003–2013. *Political Studies* 66(2): 287-305.
Baker P (2006) *Using corpora in discourse analysis.* A&C Black.
Baron A, Rayson P and Archer D (2009) Word frequency and key word statistics in corpus linguistics. *Anglistik* 20(1): 41-67.
Benzécri J-P and Benzécri F (1984) *Analyse des Correspondances: exposé élémentaire.* Dunod.
Caliński T and Harabasz J (1974) A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 3(1): 1-27.

Cardano M (2020) *Defending Qualitative Research: Design, Analysis, and Textualization.* Routledge.

Clausen SE (1998) *Applied correspondence analysis: An introduction.* Sage.

Davies DL and Bouldin DW (1979) A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence.*(2): 224-227.

Eco U (1992) Interpretation and overinterpretation. Carocci ed. Cambridge: Cambridge University Press; trad. it. Interpretazione e sovrainterpretazione, Milano, Bombiani, 2002.

Franzosi R (2008) Content analysis: Objective, systematic, and quantitative description of content. *Content analysis* 1(1): 21-49.

Greenacre M (2017) *Correspondence analysis in practice.* CRC press.

Greenacre MJ (1984) Theory and applications of correspondence analysis.

Huang A (2008) Similarity measures for text document clustering. *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand.* 9-56.

Karypis MSG, Kumar V and Steinbach M (2000) A comparison of document clustering techniques. *TextMining Workshop at KDD2000 (May 2000).*

Krippendorff K (2004) *Content Analysis: an introduction to its methodology.* Thousand Oaks, etc.: Sage.

Lancia F (2012) The logic of the T-Lab tools explained. *Retrieved 10 Jan. 21, https://mytlab.com/textscope.pdf.*

Lebart L, Salem A and Berry L (1998) *Exploring textual data.* Dordrecht, Boston: Kluwer Academic Publisher.

Mihalcea R, Corley C and Strapparava C (2006) Corpus-based and knowledge-based measures of text semantic similarity. *Aaai.* 775-780.

Rastier F, Cavazza M and Abeillé A (2002) *Semantics for descriptions: From linguistics to computer science.* Stanford Univ Center for the Study.

Rivera I (2019) Package 'RedditExtractoR', https://cran.r-project.org/web/packages/RedditExtractoR/RedditExtractoR.pdf. *accessed on Nov. 2019.*

Savaresi SM and Boley DL (2004) A comparative analysis on the bisecting K-means and the PDDP clustering algorithms. *Intelligent Data Analysis* 8(4): 345-362.

Schonhardt-Bailey C (2008) The congressional debate on partial-birth abortion: Constitutional gravitas and moral passion. *British journal of political science.* 383-410.

Sood G, Lyons K and Muschelli J (2020) Package 'tuber', https://cran.r-project.org/web/packages/tuber/tuber.pdf. *last accessed on June 2020.*

T-Lab (2021) User manual. *https://www.tlab.it/download/#documents.*

Weller SC and Romney AK (1990) *Metric Scaling: Correspondence Analysis.* London: Sage University Paper.

Appendix 2 – Alternative solutions

Tab. A The formal elements of verisimilitude of the alternative 'solutions' against CoDiRO (groups B and C)

'Solution'	Elements of verisimilitude
the Perrino's hypothesis	- Pietro Perrino is a geneticist graduated in Agricultural Sciences, former director of the Germplasm Institute of the CNR (National Research Council) in Bari; - the lexicon and some contents are inspired by quantum physics.
the Copagri 'method'	- The University of Foggia is a scientific partner, with the direct involvement of Francesco Lops, Associate professor of Plant Pathology, and Antonia Carlucci, researcher in the same scientific sector.
the lactoperoxidase 'method'	- The Lubixyl Consortium, to which the authorship is owed, claimed to be the expression of 26 universities, 34 laboratories and 15 technical and professional organizations, from 15 countries around the world; - the lactoperoxidase enzyme exists and really performs antimicrobial and antioxidant activity.
The Xiloyannis' protocol	- Cristos Xiloyannis is Full professor of Physiology of fruit species, general fruit growing and nursery techniques, University of Basilicata.
The Silecc project ("Environmentally friendly control systems against CoDiRO" project)	- Margherita D'Amico, the scientific manager, is a phytopathologist.
Dentamet® or Scortichini's 'method'	- Marco Scortichini is director of the Fruit Growing Research Unit of the Council for Agricultural Research (Crea); - Scortichini is co-author of some of the most authoritative studies on the diagnosis of CoDiRO in Puglia; - Dentamet® obtains a patent; - Dentamet® is a "biological aerosol", based on substances (zinc, copper and biocomplex citric acid) which, although ineffective on *X. fastidiosa*, are natural, as well as already accepted in the phytosanitary field.
Informational water 'method'	- The inventor, Salvatore Riamò, is a doctor; - the description mimics a lexicon that, although largely inaccurate, winks at particle physics and information theories.
Roveri's 'method'	- Norberto Roveri, retired in 2013 as a Full professor of General Chemistry at the University of Bologna; - Roveri directs the Chemical Center srl, a spin off of the University of Bologna active in the microparticle sector; - the product based on calcium phosphate microparticles obtains a patent.
BiCC project (Bio-Contrast to CoDIRO project)	- Giusto Giovannetti, the scientific manager, is a biologist; - Giusto Giovannetti is the scientific director of CCS srl of Aosta, a company that has been active for 30 years in the

research and production of micro-organisms for agriculture and for environmental remediation from diffuse pollution;
- CCS srl is included in the national research register;
- the project obtains funding from the Puglia local government.