



COESO

connecting research and society

COLLABORATIVE ENGAGEMENT ON SOCIETAL ISSUES

WP5 - Cooperation Quality Assessment

Report on activism and science within Citizen Science

22.12.2023



This deliverable is licensed under a Creative Commons Attribution 4.0 International Licence.



COESO has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020) SwafS-27-2020 - Hands-on citizen science and frugal innovation, under Grant Agreement No.101006325

The content of this publication is the sole responsibility of the COESO consortium and can in no way be taken to reflect the views of the European Commission. The European Commission is not responsible for any use that may be made of the information it contains.

Deliverable 5.3

Report on activism and science within Citizen Science

Grant Agreement number	: 101006325
Project acronym	: COESO
Project title	: Collaborative Engagement on Societal Issues
Funding Scheme	: H2020-EU.5. - SCIENCE WITH AND FOR SOCIETY
Topic	: SwafS-27-2020 - Hands-on citizen science and frugal innovation
Project's coordinator Organization	: EHES-OpenEdition
E-mail address	: pierre.mounier@openedition.org , alessia.smaniotto@openedition.org
Website	: https://coeso.hypotheses.org
WP and tasks contributing	: WP5 - Cooperation quality assessment
WP leader	: FN-SP-CEE
Task leader	: FN-SP-CEE
Due date	: 31.12.2023
Delivery date	: 22.12.2023
Authors	: Jessica Pidoux, Dominique Boullier, Yuma Ando (FN-SP-CEE)
Reviewer	: Maxim Kupreyev (OPERAS)

Contents

Executive Summary	4
1. Introduction	4
2. Theoretical Framework: Orders of Worth	6
Computational implementation of orders of worth	9
Justification of theoretical framework within citizen science practices	11
3. Data Collection	12
Guidelines	12
Second round of pilots No. 6, 7, 8, 9, 10	13
Personal data protection	14
Low adoption of guidelines	14
Text pre-processing	15
3.1. Datasets	17
On-going activities of pilots: Blog hypotheses' outputs	18
On-going oral and written communication: Mails and meetings	18
Mails	18
Meetings	19
4. Feasibility Study: Algorithms' development	20
4.1. First technique: Word2Vec	20
4.1.1. Dictionaries for NLP: Common principles of orders of worth	23
Word2vec's Results: Non-significant	23
4.2. Second technique: Sentence-Transformers and zero-shot classification	24
Sentence-Transformers and zero-shot Results: Non-significant	25
5. Results Orders of Worth Pilot 7	27
5.1. Manual identification of conflict situations and classification of orders of worth	27
5.2. Limits of Orders of Worth's technical implementation	28
Conclusion: Lessons Learned	30
References	32
Annex	33
1. Data Collection Guidelines for COESO's Pilot Projects	33
List of tables	38

Executive Summary

The final deliverable of Work Package 5 (WP5) Cooperation Analytics within the COESO project is centred on Task 5.3, "Observation and analysis of conflicts of justification between activism and research activities", employing a theoretical framework and computational methods designed to understand discourses within citizen science practices of selected pilot projects, incorporating textual analysis. It introduces the theoretical framework of Boltanski and Thévenot [1991] (2006) known as "On justification: economies of worth." This framework, encompassing seven conflicting "orders of worth," offers a lens to understand and analyse the justification processes in social life disputes. Given the diversity of stakeholders in citizen science projects, conflicts between different perspectives and values are valuable for defining citizen science conventions for the sake of accountability. The methodology employed involves a combination of semi-automation methods and qualitative analysis for textual data, aligning it with the theoretical framework. Outputs include dictionaries of keywords signalling different orders of worth, the results from a feasibility study on two algorithms tested on collected data, and insights into the limitations and opportunities of automated and qualitative analyses. The lessons learnt emphasise the requirement for future research to detect and understand citizen science's moments of conflicts for their accountability and facilitating reflexivity among stakeholders about their scientific practices.

1. Introduction

The final deliverable of WP5 Cooperation Analytics of the project COESO comprises task "T5.3 Observation and analysis of conflicts of justification between activism and research activities" with a theoretical framework and computational methods that were built to analyse the discourses between activism and scientific practice arising during practices of selected pilot projects, including textual analysis. The task is part of the general goal of WP5 to develop cooperation analytics, i.e. a set of indicators measuring cooperation practices between members of pilot projects.

More specifically, "D5.3 Final report on activism and science within Citizen Science" presents *On justification: economies of worth*, a theoretical framework from Boltanski and Thévenot [1991] (2006) that we applied to citizen science practices occurring during the COESO project. According to Boltanski and Thévenot ([1991] 2006) the theory of justification comprises a model of seven conflicting "orders of worth" (*grandeurs* in French), six initially conceptualised i.e., inspired, domestic, market, industrial, opinion, civic, and an additional one, i.e. project, developed later by Boltanski and Chiappello (1999). These orders present a

plurality of modes of justification in social life disputes that are made explicit. At the core of the justification theory is how social actors explicate the grounds of their choices, their behaviour and propositions; the action that is “just” to them according to a given order of worth, which has its proper virtues and principles to coordinate the course of action.

Considering the diversity of the stakeholders in citizen science projects, conflicts between different standpoints and values are not a rare situation but a rather normal one. The need to elicit these different principles can be traced during each discourse activity of the on-going cooperation. Detecting these moments and their expressions should help participants gain some reflexivity on the foundations of the project and adapt their behaviour based on repertoires of expression of conflicts using typical registries of language.

We combined semi-automation methods and qualitative analysis to analyse the textual data with respect to the theoretical framework. Our outputs include first, two dictionaries with a list of keywords and expressions that signal the different orders of worth. Second, the results of a feasibility study conducted on two algorithms that were tested on the collected data. They can be implemented in further research with a consequent dataset of citizen science practices. Finally, the report includes the limitations of the automated and quantitative analysis, as well as the opportunities offered by a qualitative analysis when the text material is analysed by experts. A qualitative analysis is more accurate for understanding arguments, identifying situations of conflicts, and moments of justifications between orders, according to a theoretical framework adopted.

In the following, section 2 presents our theoretical framework, as well as its application to citizen science practices and computational measurements. Section 3 presents the data collection process with the different Pilots’ datasets and the selected ones that were used. It also includes the limitations that were faced during the collection. Section 4 covers the feasibility study for algorithms’ development: an explanation of the techniques used and the results obtained. The results of the two tests are presented with their respective challenges faced. It also contains two dictionaries in French and English. Section 5 presents the results of the orders of worth’s qualitative analysis for pilot 7. The conclusion is focused on the lessons learned during the realisation of task 5.3 to open up perspectives in the use of natural language processing based on a sociological theory in citizen science.

2. Theoretical Framework: Orders of Worth

Within the COESO project, WP5 cooperation analytics seeks to understand how new conventions of citizen science practices are being developed by the actors themselves. Conventions are defined by Boltanski and Thévenot [1991] (2006) as common benchmarks that enable actors to participate in a collective action. These common benchmarks, that can be explicit or implicit, enable actors to understand each other and to cooperate based on principles and values that guide the coordination of their actions without the necessity of verifying such conventions. Making these conventions explicit are crucial for making scientific practices robust and verifiable.

A theoretical framework to extract the elicitation process of novel conventions within their process of being built between civil society and academia is Boltanski and Thévenot's [1991] (2006) *On justification: economies of worth*. They drew within their pragmatic sociology, a theory of justification with a model of six conflicting "orders of worth" (grandeur), i.e., inspired, domestic, market, industrial, opinion, and civic. Some years later, Boltanski and Chiappello (1999) added one more order of worth, the one of "project". These orders present a plurality of modes of justification in social life disputes. At the core of the justification theory is how social actors make the grounds of their choices explicit, their behaviour and propositions; the action that is "just" to them according to a given order of worth. Through their justification, it is possible to have access to the social actors' discourse on which new agreements can be reached, the way they evaluate and orient actions for any purpose. Only some specific moments of interaction will require these justification procedures, when the "natural" course of action is not shared anymore, or when it is formally challenged through "tests" or "critiques", i.e. situations of disputes with explicit arguments. In these situations, every tacit natural social signal needs to be expressed, assessed and negotiated through "compromises" (beyond the more frequent tacit "arrangements"). Social situations are usually pre-arranged and consist of many features including objects, locations, rituals, data, etc. Indeed, actors do not engage in actions in the same way if they are at home, at work, in a church or in a park and each of them will orient members of the society towards some supposedly shared normal behaviour and values, where some are greater and more conventional than others.

The "orders of worth" are based on different principles and criteria that actors use to evaluate the worth or importance of things, such as achievement, solidarity, or distinction. Their theory offers a way to understand the diversity of practices in social life and how they can conflict with one another. It is also relevant to understand negotiation and compromise when resolving conflicts between different orders of worth. This relates to forms of cooperation in linguistic

material, as an extension of WP5's cooperation analytics. It is important to note that for such a justification theory, the main reference to define orders is discursive; what actors say in conflicting situations. Indeed, the authors defined orders from a corpora of manuals that can be considered typical of the best practices in a given environment (legal document, manual for behaving in formal dinners, etc.). They systematically extracted the terms used to demonstrate the value of this specific social world, including the ones mentioning the objects supposed to be natural in these situations: for instance, at the moment a chronometer is displayed in a room, the references and the values to be applied by the participant change radically to become part of the "industrial" order of worth in which is great what is efficient and controlled.

To understand the orders of worth, initially also called "social worlds", the authors refer to the notion of "worth" as to understand the virtues, principles and personal criteria in given situations where a social actor can be found as: worth something, worth by somebody, be worthier than... less worthy than..., or be the worthiest.

The first order is *inspired*, where persons may be more or less worthy inasmuch as they are all capable of experiencing inspiration. Actions are here justified based on inspiration, creativity, and aesthetic criteria. For instance, an artist creating a unique and unconventional piece of artwork can justify his actions as driven by personal inspiration and a desire to express creativity.

The second order is *domestic*, which emphasises the importance of familiarity, routine, and the preservation of the established social order. Actions and choices are justified based on their alignment with traditional values, customs, and stability. It's related to oneself, one's position and body, reducing uncertainty and errors of misidentification in personal encounters. Here justification is grounded in principles of family and care, tradition and hierarchy. For instance, a young person sacrificing personal career ambitions to take care of an older family member.

The third order is *civic*, that is focused on the individuals forming a collective, under legal procedures that allow them to justify their representativity. Here actions are justified by reference to principles of civic duty, the common good, and public service. Civic worth is associated with principles of justice, equality, and public service. For instance, a group of volunteers participating in a cleanup initiative, motivated by a sense of duty in their neighbourhood and a commitment to enhancing the cleanliness of public space.

The fourth order is *market*, where individuals are compelled by their desires to acquire objects, particularly rare goods with transferable ownership, which involves a competitive aspect. For instance, rich persons live the high life to own what they

want like luxury items, which they know how to sell and which is expressed by their success.

The fifth order is *industrial*, which is based on efficiency, productivity and capacity to ensure organisational operations; one can find academia here as it is not limited to industry. For instance, a manager in a company investing in advanced technology for reducing costs in production processes, guided by principles of time optimization, human labour minimization and scientific progress.

The sixth order is *opinion*, previously called “fame”, which establishes equivalence, and the worth of each entity (thing or person) depends on the opinion of others, that is looking to improve its reputation through an extension of its audience. What is great is what is renowned, famous, despite all other weaknesses in other orders of worth (one can be famous but considered as incompetent, not representative, and so on). For instance, a public figure or celebrity adjusting their behaviour to align with public expectations and maintain a positive public image, emphasising the importance of public opinion and recognition, especially through media exposure.

Table 1: Six orders of worth

	Inspired	Domestic	Civic	Opinion	Market	Industrial
Mode of evaluation (worth)	Grace, non conformity, creativeness	Esteem, reputation	Collective interest	Renown	Price	Productivity, efficiency
Format of relevant information	Emotional	Oral, exemplary, anecdotal	Formal, official	Semiotic	Monetary	Measurable, criteria, statistics
Elementary relation	Passion	Trust	Solidarity	Recognition	Exchange	Functional link
Human qualification	Creativity	Authority	Equality	Celebrity	Desire, purchasing power	Professional competency, expertise

Source: L. Boltanski and L. Thévenot, 1999, "The sociology of critical capacity", *European Journal of Social Theory*, vol.2, n°3, special issue "Contemporary French Social Theory", (August).

Additionally, as an extension of our main framework, we define the seventh order of worth, i.e. *project* (or *network*) according to *The New Spirit of Capitalism* of Boltanski and Chiapello (2011). This order is associated with a managerial and organisational logic prevalent in contemporary capitalism. The project is a set of

values and principles that guide economic and organisational activities in a distinct way from industrial order.

For instance, one can find in this order a social actor initiating and managing projects, valuing innovation, adaptability, and flexibility in responding to changing circumstances. It is also about promoting collaboration and teamwork as essential for project success in a situation where the project is seen as a source of personal fulfilment and a means for individuals to express their talents and aspirations.

Now the core of the theory is how social actors justify their choices in moments of dispute between multiple orders of worth interacting to reach a possible agreement. For example, the industrial order is based on the principles of efficiency and productivity. In this order, individuals are valued for their ability to display behaviours based on speed, performance, high quality, compliance with metrics of various kinds, etc.. The domestic order is focused on the care and nurturing of family and community. These two orders of worth can be seen as potentially problematic when interacting, as it can lead to a focus on successful performance indicators at the expense of other values and social relationships.

Computational implementation of orders of worth

From this theoretical framework, we examine the way these orders can be identified and analysed within linguistic materials using computational methods such as natural language processing, that uses statistics, like probabilistic models, for interpreting linguistic expressions with computer science techniques.

The technical aim within task 5.3 comprises two tasks: identifying if a given (textual) situation belongs to a given order; identifying a conflict situation to find trends in discussions of pilots' projects through keywords and expressions. These "conflict situations" can be found in communication practices between project members about the general orientation and priorities of the project or about more local situations where a critique is formulated about a specific behaviour that is evaluated and disqualified from a specific order of worth. In this context, the term "conflict situation" means a communicative moment of dispute, sometimes very intense, sometimes very soft, but anyway making explicit the opposite valuation of a situation or a behaviour, or a person.

For illustration purposes, the two following discussions present a fictional situation of conflicting orders, between two supposedly members of a pilot project in citizen science that we could use as textual data for automatic processing.

Situation 1

Personne 1 : Dis donc, je voudrais savoir quand est-ce que les publications du projet seront mises en ligne ?

Personne 2 : Tu parles de quelles publications ?

Personne 1 : Toutes en général.

Personne 2 : Parce que Clara, elle a ses impératifs de publication pour sa revue et elle vise un public large, donc ça demande une réécriture pour toucher l'opinion et non seulement fournir un compte-rendu. Il doit y avoir des visuels tirés des événements et des choix de présentation beaucoup plus attractifs, ce qui va lui demander plus de temps.

Personne 2 : Mais ça m'étonne que tu dises ça, parce qu'on avait prévu de garder ces contenus pour vendre une plaquette, parce que c'est prévu dans notre budget, on doit générer des revenus avec ces contenus audiovisuels ciblés, il nous faut des acheteurs solvables et donc il n'est pas question de les mettre à disposition juste pour la réputation du projet. Je te rappelle qu'il y a de la concurrence sur ce domaine et il faut qu'on prenne une part de marché.

Personne 1 : Mais trouver un large public et gagner en visibilité c'est une condition de départ pour réussir à vendre. Là, tu fais une erreur de priorité. Tout le monde sera très content que ces contenus soient partagés et célèbres et cela donnera de la reconnaissance à chacun et à tout le groupe. Même si ça retarde un peu la commercialisation, ce n'est pas très grave.

This discussion displays an opposition between seeking commercial performance (order of market worth) and pursuing a global reputation through media exposure (order of opinion worth). No one is intrinsically right or wrong, there is a conflict of priorities, of strategies that pushes the members to explicit what is worth for them. It may end into a split, a shift in the project orientation, or a compromise but this moment of dispute is typical of a justification conflict that can be found in many projects. All terms can be labelled referring to a specific order of worth and this detection can be automatized. However, this is a fictional controversy with many explicit expressions referring to two clearly different orders of worth. Real situations are much more fuzzy and based on references that are not easy to elicit, traced and computed.

Situation 2

Personne 1 : Nous devons prendre une décision concernant le budget de notre projet de bénévolat en faveur des sans-abri.

Personne 2 : Oui, il est important de s'assurer que nous avons suffisamment d'argent pour acheter des couvertures et des vêtements chauds pour les personnes dans le besoin. La compassion et la générosité sont des valeurs importantes pour notre organisation.

Personne 1 : Oui, mais nous devons également être conscients de nos propres limites financières. Nous ne voulons pas mettre en péril la viabilité de notre organisation en dépensant trop d'argent. L'efficacité et la responsabilité financière sont également des valeurs importantes pour nous.

Personne 2 : Je comprends ce que tu dis, mais je pense qu'il est plus important de s'assurer que les personnes dans le besoin ont ce dont elles ont besoin plutôt que de protéger notre budget. La solidarité et l'aide à autrui sont des valeurs encore plus importantes pour notre organisation.

This discussion, a rather formal and abstract one, not so realistic was produced in order to conduct a first test of our algorithms. It displayed a conflict between domestic order of worth (solidarity, care) and budget requirements (market order of worth).

From a computational perspective, this detection is a difficult issue as we need to define the conflict communicative situations in on-going practices of pilot projects, which are not usually made explicit in written communications. They often happen in spontaneous and informal discussions.

Justification of theoretical framework within citizen science practices

In citizen science projects, it is commonly supposed that all participants would share 2 orders of worth at least as a common ground for understanding: the industrial one (we produce science with all its requirements of argumentation, documentation, control and publication) and the civic one (citizens contribute to scientific production of knowledge and this extension of scale is also an extension of the values that must guide the projects, i.e. the common good and the interests of large communities, including their explicit recognition as valuable contributors to knowledge). However, it remains unclear how the balance, the compromise between these two principles and sets of values will be implemented. Some will consider that citizens must deliver some information from their observations and that this is a valuable output for everybody by the extension of resources and data that is obtained (e.g. crowdsourcing). Some others would consider that the citizens should be the ones to orient the main goals of a project, for the benefit of some communities, and that scientific knowledge will be of interest only if it delivers these outputs for citizens. And many other mix orientations can be found.

What is important, is that since a project is a collective learning experience, these starting assumptions can be and often must be revised. This is where some disputes become very profitable for the cooperation, because the implicit

background framework of all members may need to be refreshed, reestablished, consolidated, or deeply reconsidered and revised.

Citizen science generates such a cultural challenge between different standpoints within a project that participants must be aware of this, and consider it as a normal situation, where some moments of dispute will occur. Being prepared for it by the detection of emerging conflicts of orders of worth can improve the self assessment of a project and the negotiation of reorientation, if needed.

The linguistic material produced during any cooperation process is the only source to trace that and detect these moments. Some projects may be more “talkative” than others or more prone to self assessment, or more ready to share their data of their everyday conversations with social scientists and algorithms. These are the conditions of feasibility that we had to experience and go through which may explain the limitations of our results.

3. Data Collection

Data collection involved considering the multiple platforms used by pilots. The diversity of formats and variable quantity of data sources per pilot depended on the platforms they used, e.g. Gmail, institutional email, Whatsapp, Zoom, hand written notes, Sharedocs. This was a main lesson learned during the first phase of our project as explained in the deliverable “Report on Test and Final Development of the Cooperation Analytics” (COESO D.5.2), URL: <https://zenodo.org/records/6787834>.

Guidelines

The diversity of data sources and platforms used posed challenges for data processing and analysis. Therefore, to standardise the data collection across pilots, we developed “Data Collection Guidelines for COESO’s Pilot Projects” (see Annex 2). The guidelines allowed us to ensure a minimum of feasibility for data processing.

The guidelines included:

1. recording meetings systematically when they were took place online,
2. using a template provided for note-taking during the meetings,
3. scheduling meetings in an online calendar with details of the agenda and members invited,
4. mails specifically categorised as COESO-related communications,
5. discussions related to the project to be migrated to VERA’s Mattermost,
6. profile page created proper to the cooperation analytics,

7. documents related to the meetings stored in a common repository in Sharedocs <https://sharedocs.huma-num.fr> via the pilots' account at <https://humanid.huma-num.fr>

Additionally, we included the posts produced by pilots in COESO's blog Hypotheses as we detail later in the dataset section below.

Second round of pilots No. 6, 7, 8, 9, 10

The data collection guidelines concerned the second round of pilots selected in an open call of the COESO project. 5 new pilot projects were chosen during a process that ran from November 2021 until May 2022, see <https://coeso.hypotheses.org/about/open-call-faq>. For more details see their specific reports of activities, and blog <https://coeso.hypotheses.org/pilots>. We present them here briefly for the sake of overviewing their goals and the scope within data for WP5 was collected.

Pilot 6: Digital Mapping with Disabled Citizens (DiMDiCi): Engaging a group of people with disabilities in the co-design of an inclusive digital collaborative mapping tool in Herne, Germany.

COESO Partners: Hochschule für Gesundheit, Department of Community Health (HS Gesundheit – Germany), University of Twente, Faculty of Geo-Information Science and Earth Observation (UT-ITC – The Netherlands), Diakonische Stiftung Wittekindshof (Germany), Municipal Administration of the city of Herne, Department of Health (Germany)

Location: Herne, Germany

Pilot 7: Ageing in a Caring Community (AGORAge): Fostering the societal value of ageing people in an Italian community.

COESO Partners: Rovira i Virgili University (Spain): DAFITS (Department of Anthropology, Philosophy and Social Work) and MARC (Medical Anthropology Research Center), ISRAA- Istituto per Servizi di Ricovero e Assistenza agli Anziani (older adults public service provider, Italy)

Location: Italy

Pilot 8: Women Water Watch (wWw): Mapping water quality from the river to the glass in Tanzania.

COESO Partners: University of Antwerp, Institute of Development Policy (IOB – Belgium), Aqua Farms Organisation (AFO – Tanzania)

Location: Tanzania

Pilot 9: Playful Futures: Sci-fi Living Action Role Play (LARP) ethnography for Mediterranean coastal communities. Raising awareness about climate change through gaming activities in Croatian coastal communities.

COESO Partners: Edgeryders OÜ (Estonia), TANTlab – University of Aalborg (Denmark), Culture Hub Croatia (CHC) (Croatia)
Location: Croatia

Pilot 10: LUNCH-BOX-MONITOR: Insight into the nutritional quality of school lunch-boxes to assess food insecurity among primary school children.

COESO Partners: Ghent University (Belgium), Let Us (Belgium), Rikolto Belgium, Flemish Institute of Healthy Living (Belgium).
Location: Belgium.

Personal data protection

The guidelines ensured data protection following our data management plan available within the overall COESO DMP (D.1.2, last version delivered at M36, URL: <https://zenodo.org/records/6787951>). The ethical and legal aspects are in alignment with Sciences Po's and COESO project privacy requirements. The main feature of these guidelines was that researchers did not access the pilots' data directly. Pilots were informed of the data processing purpose and research in advance. Then, they were asked to provide selected data types from specific sources themselves following the guidelines. We minimised this way the personal data collected that was later anonymised for analysis.

Low adoption of guidelines

The adoption of the guidelines by pilots was low as shown in Table 2, the data sources that were exploited vary across pilots. The most complete dataset was provided by one Pilot No. 7 with 5/6 sources, followed by three Pilots No. 8, 9, 10 with 4/6 sources, and Pilot No. 6 with 2/6 sources.

Table 2. Data collected per pilot based on the different sources exploited in the guidelines

Pilot No.	Meeting minutes	Meeting recorded	Emails	Profile page	Calendar	Hypotheses
Pilot 6	no	no	no	yes	no	yes
Pilot 7	yes	yes	yes	yes	no	yes
Pilot 8	yes	no	yes	yes	no	yes
Pilot 9	yes	no	yes	yes	no	yes

Pilot 10	yes	no	yes	yes	no	yes
----------	-----	----	-----	-----	----	-----

The low adoption of these guidelines had three reasons.

The first reason was that this required additional work for pilots as new protocols had to be integrated into their daily practices that were initially not consistent through time, nor made in a digital support by default, for instance when some team members in Pilot 8 were doing field work without any digital dispositif.

The second reason was that these guidelines required separating COESO related data with other types of data produced in their occupation. Therefore, this division of data types and transmission to WP5 demanded additional work from pilots.

The third reason, that only concerned one pilot No. 6, was that they were collecting sensitive data. Therefore, to ensure their data subjects' privacy and a high level of compliance with their institution, they decided to not share their data with COESO, even though it was originally planned.

Overall it was a challenge to find a balance between standardising practices across pilots that would contribute to create conventions in citizen science, and allow them to preserve the diversity that characterises each pilot, since that was one of the main sources for selecting them in the open call.

Text pre-processing

Preprocessing and cleaning datasets are crucial steps in preparing data, both structured and unstructured, for natural language processing (NLP) tasks. These steps help ensure that the data is in a suitable format for analysis and modelling. Data preprocessing involves a combination of manual and automated techniques that we applied in our corpora as explained in detail in our previous deliverable "Report on Test and Final Development of the Cooperation Analytics" (COESO D.5.2), URL: <https://zenodo.org/records/6787834>. While automation can handle repetitive and standardised tasks efficiently, manual intervention is often necessary for tasks that require domain expertise, nuanced decision-making, or handling specific nuances in the dataset.

There were six main tasks of pre-processing:

1. Scraping text from Pilots' hypotheses blog
<https://coeso.hypotheses.org/category/blogposts/pilots>.

This textual data was relevant as an output of the pilots' cooperation practices in continuous production during their projects. Not all pilots planned scientific publications or have not published articles during the project. Therefore, the blog hypotheses developed within COESO was a main source of keeping track of the on-going outputs of the pilots' activities as every pilot published posts after every main activity they conducted.

The code for data extraction and analysis is available at:

[https://github.com/yuma-ando/COESO Cooperation Analytics/blob/main/Notebook/Script for scraping.ipynb](https://github.com/yuma-ando/COESO_Cooperation_Analytics/blob/main/Notebook/Script_for_scraping.ipynb)

2. Text extraction from COESO WP5's generic email addresses created for pilots.

Following our data collection guidelines, there were five email addresses, one for each pilot, that were created to collect dedicated data about cooperation analytics. These were:

coesopilot6@sciencespo.fr

coesopilot7@sciencespo.fr

coesopilot8@sciencespo.fr

coesopilot9@sciencespo.fr

coesopilot10@sciencespo.fr

This textual data included on-going communication practices, email exchanges, between members of a pilot project. The extraction was done from those email addresses to which WP5 had access. The extraction required additionally to separate textual data from metadata (i.e. senders, C.C, timestamp).

The code for extraction is available at:

[https://github.com/yuma-ando/COESO Cooperation Analytics/blob/main/Notebook/Tool for extracting filtering and cleaning GMAIL data.ipynb](https://github.com/yuma-ando/COESO_Cooperation_Analytics/blob/main/Notebook/Tool_for_extracting_filtering_and_cleaning_GMAIL_data.ipynb)

3. Standardisation of textual data extracted from different sources.

This required data automatic cleaning by removing unnecessary special characters, or symbols, converting it to lowercase, removing stop words, and stemming or lemmatizing words. Furthermore, it also required manual cleaning including the elimination of greetings, signatures at the end of the email, as well as structuring of separated sentences in units after commas and dots for analysis. There were two iterations for this process. The cleaning included also identifying and removing duplicate records from the dataset as well as identifying and handling outliers that may affect the quality of the data.

An example of the standardisation code is available at:

[https://github.com/yuma-ando/COESO Cooperation Analytics/blob/main/Notebook/Changes_in_work_formalisation_and_Moments_of_work_revision.ipynb](https://github.com/yuma-ando/COESO_Cooperation_Analytics/blob/main/Notebook/Changes_in_work_formalisation_and_Moments_of_work_revision.ipynb)

For the detailed description of these three first processes, see sections “Data structure study” and “Extraction and analysis methods” in deliverable “Report on Test and Final Development of the Cooperation Analytics” (COESO D.5.2), URL: <https://zenodo.org/records/6787834>

4. Conversion of video files.

For the recorded meetings it was required to convert the format from MP4 to MP3 for transcribing the audio data and separate it from the video data that was not going to be used.

5. Automated transcription of videos.

The recorded meetings selected for analysis were from Pilot 7 (who had a more complete and consistent recording of their meetings, see datasets section). The audio files of videos were transcribed from format MP3 with Whisper (Openai). “Whisper is an automatic speech recognition (ASR) system trained on 680,000 hours of multilingual and multitask supervised data collected from the web”. Source: <https://openai.com/research/whisper>. A speech-to-text task was accomplished. In order to be GDPR compliant, an open source version of Whisper was locally installed on the team member’s computer and the meeting data was locally transcribed into a text file. Whisper was used as a command line application.

6. Automated translation of videos.

A translation of audio files’ transcription was coupled with the speech-to-text task previously done, namely the output text file was automatically translated from Italian to English with Google Translate via API. Source: <https://cloud.google.com/translate/docs/reference/rest/>. The idea was to analyse its quality to be able to translate the different languages used by pilots in future development. Pilots were communicating in eight languages; spanish, english, french, portuguese, italian, german, africans, dutch. However, only few did it punctually, e.g. during COVID (first round of pilots) and they were not doing it systematically. That is why we focused on pilot 7.

A qualitative review revealed however that the nature of conflict is often expressed in a subtle way, which requires text data as close as possible to the original audio file. Given a loss in data quality during the speech-to-text task, the transcription in the original language was finally preferred.

The translation task was based on the same script as the one used to standardise the textual data. (See task 3 Standardisation of textual data extracted from different sources).

3.1. Datasets

There were two main datasets built from our data collection process of pilots. Textual data from hypotheses blog, and textual data from emails and meetings.

On-going activities of pilots: Blog hypotheses' outputs

The blog hypotheses is a useful textual corpora where each pilot has a page. Pilots continuously published posts throughout their activities that could be linked to the orders of worths. Although the corpora consists of pilots explaining their practices once they are finished, and not their ongoing practices, we found their continuous publication to be a starting point for translating the theoretical framework into computational models.

The five blogs analysed were the following:

1. Pilot 6 | Digital Mapping with Disabled Citizens in Germany | <https://dimdici.hypotheses.org/>
2. Pilot 7 | Ageing in a Caring Community | <https://agorage.hypotheses.org/>
3. Pilot 8 | Women Water Watch | <https://waterwatch.hypotheses.org/>
4. Pilot 9 | Playful Futures: sci-fi Live Action Role Play ethnography for Mediterranean coastal communities | <https://playfutures.hypotheses.org/>
5. Pilot 10 | Lunch-Box-Monitor | <https://lunchbox.hypotheses.org/>

A first sample test was built on two pilots' blogs from COESO's first round of pilots:

Pilot 2 Dancing Philosophy <https://dansophie.hypotheses.org/>

Pilot 3 Social Evolution & Mafia <https://usbc.hypotheses.org/>

22 posts of blogs were extracted and analysed from those two blogs, this included French and English transcriptions as they were not all published in the same language. The translation allowed us to have multiple datasets to try the models' accuracy in different languages.

On-going oral and written communication: Mails and meetings

Mails

Four pilots provided their email exchanges for a dataset composed as follows:

Table 3. Mail dataset

Pilot No.	Qty.	Period
6	0	N/A

7	47	Aug.2022-Feb.2023
8	35	Jul.2022-Nov.2022
9	74	Jul.2022-Nov.2022
10	62	Oct.2022-Jan.2023
COESO WP5 members	196	Mar.-Jul. 2022

Preliminary tests were conducted by creating a dataset of email exchanges between two members of COESO WP5.

In addition to the standardisation issue raised above, the nature of the message exchange involved two practical challenges for the text analysis of this study. Firstly, the majority of exchanges are rather factual messages, including the information about the meeting date and other greetings. We seldomly observe a substantial discussion concerning the orientation of the project where any sort of justification is explicitly argued. Secondly, given the temporal separation of exchanges, a confrontation of arguments is difficult to identify. Since a message tends to convey a set of information at the same time, there is no continuous discussion on a specific subject. With the presence of additional information and sentences on other topics, a conflictual moment is hardly spotted in text analysis. For those reasons, the text data from mails was not the primary source of data used in this study.

Meetings

Three pilots provided meeting minutes, and one pilot provided meeting recordings for a dataset composed as follows:

Table 4. Meetings dataset

Pilot No.	Qty.	Period
6	0	N/A
7	8 mp4 videos, 4 Word documents, 17 PDF documents	Sept. 2022-Apr. 2023
8	7 Word documents	Jun.-Nov. 2022
9	10 Word documents	Jul.-Nov.-2022
10	0	N/A

The meeting meetings presented three main problems for analysis:

1. Insufficient data in terms of size and period

2. Text written as summaries of conversations
3. Speakers were not identified systematically across sources

We selected Pilot 7 meeting recordings as the most useful resource for analysing orders of worth. The videos provide real-time interactions, explicit discussions, within a longer period of time than the meeting minutes provided by other pilots.

Therefore, Pilot 7 was selected as our experimental case for the analysis of orders of worth. The following sections of the deliverables are focused on this pilot.

4. Feasibility Study: Algorithms' development

For analysing text corpora based on the theoretical framework defined we conducted a feasibility study for the development of algorithms testing two techniques: word2vec and a combination of semantic similarity and zero-shot classification.

The goal was to develop an algorithm which can detect potential moments of conflicts (and how those conflicts are based on a confrontation of different orders of justification) in a continuous process, so that the algorithm can monitor the interactions between numerous members of a collaborative project, using different sources of information (written documents, blogs, deliverables, meetings, etc.). This goes beyond Boltanski and Thévenot's original research - their sources were "pre-categorized" through the background knowledge that a certain text belongs to a certain context.

The first iteration was based on Word2Vec algorithms where we used a dataset of emails, of two COESO members for the first development, and of all pilots to conduct further tests on their practices. These algorithms main function was identifying and matching keywords with orders of worth. In other words, the idea was to extract semantic fields from the given texts and compare them to the existing keywords, which belong to the "orders of worth". However, after a qualitative analysis of the results, it was identified that the algorithms did not provide any relevant result by retaining keywords for potential conflict identification according to the sense of the communication taking place. Yet, we produced a list of keywords and signals of expressions from the theoretical framework for future research.

The second iteration was based on a combination of two tasks: firstly, semantic similarity based on sentence-transformers model to refine the identification of keywords, and secondly, a zero-shot classification of sentiments expressed in the text. This variation of sentiment analysis served only in a limited manner to the

analysis as a preliminary identification of moments of conflicts according to more or less negative tones in sentences.

4.1. First technique: Word2Vec

The first algorithms tested were based on the word embedding through word2vec technique. They were tested on different text corpora: emails from COESO members, blog posts produced by pilots, minutes and pilots' emails.

Word2vec¹ “is a technique for natural language processing (NLP) published in 2013. The word2vec algorithm uses a neural network model to learn word associations from a large corpus of text. Once trained, such a model can detect synonymous words or suggest additional words for a partial sentence. As the name implies, word2vec represents each distinct word with a particular list of numbers called a vector. The vectors are chosen carefully such that they capture the semantic and syntactic qualities of words; as such, a simple mathematical function (cosine similarity) can indicate the level of semantic similarity between the words represented by those vectors”.

A first approach taken was *semantic search and matching* for detecting the different orders in text according to the literature and classifying them within specific sentences in corpora.

The methods tested included computing similarities based on a specific metric called cosine similarity, in order to find similarities between different word embeddings, while taking into account relative importance(weight) of different terms through TF-IDF. The word embedding was calculated based on a pre-trained dataset “glove-wiki-gigaword-50“, which is available in Gensim Python library (see below for a more detailed description).

The process consisted in taking all data from email users as input, pre-processing it and then applying semantic search. It searches for those parts of the text in the input data that are most semantically similar to the vocabularies of the 7 orders of worth described above. This approach is interesting in view of omitting one of the most difficult stages of processing for a computing system, namely the selection of conflict situations. However, results are not always accurate and it requires manual validation by a human. We especially noticed that the distinction of different logics of justification is not clearly established in the output, where the same sentence can be pointed out as semantically similar to various types of justifications.

¹ Source: <https://en.wikipedia.org/wiki/Word2vec>

To provide a more detailed description, four measures were applied using Gensim. Gensim² “is a Python library for topic modelling, document indexing and similarity retrieval with large corpora” in natural language processing (NLP).

The first measure applied was "Term Frequency-Inverse Document Frequency Model." It is a statistical measure used in information retrieval and text mining to evaluate the importance of a word in a document relative to a collection of documents. The TfidfModel in Gensim is used to transform a bag-of-words representation of a document into a TF-IDF weighted representation.

The second measure applied was the WordEmbeddingSimilarityIndex class from the gensim.similarities module. Word embeddings are dense vector representations of words in a continuous vector space. They capture semantic relationships between words. The WordEmbeddingSimilarityIndex in Gensim is a similarity index that computes the similarity between documents based on their word embeddings.

The third measure applied was the SparseTermSimilarityMatrix class from the gensim.similarities module. A term similarity matrix measures the similarity between terms (words or phrases) based on their co-occurrence patterns in a corpus. The SparseTermSimilarityMatrix in Gensim allows for efficient computation of term similarity using sparse matrices, which can save memory and computational resources.

The fourth measure applied was SoftCosineSimilarity class from the gensim.similarities module. Soft Cosine Similarity is a metric used to calculate the similarity between documents, taking into account semantic similarity between words. It uses the cosine similarity metric after applying a technique called soft cosine. Soft Cosine Similarity is especially useful when dealing with text data where synonyms or related terms are important for similarity measurement.

The code is available at:

[https://github.com/yuma-ando/COESO_Cooperation_Analytics/blob/main/Notebook/Orders of Worth Semantic search and Matching.ipynb](https://github.com/yuma-ando/COESO_Cooperation_Analytics/blob/main/Notebook/Orders%20of%20Worth%20Semantic%20search%20and%20Matching.ipynb)

² Source: <https://pypi.org/project/gensim/>

4.1.1. Dictionaries for NLP: Common principles of orders of worth

The first test was using two dictionaries in French and English from the theoretical framework for training NLP models. It is a list of keywords corresponding to the orders of worth, which allows the theory to be adaptable to data analysis. Given its length, the list can be found in Table 5 (see list of tables). It compiles the common principles of the seven orders of worth from our theoretical framework. These are the principles that guide the data analysis and computational methods in task 5.3 in order to extend the repertoires and to detect the various justifications within pragmatic context of cooperation. They are presented in French from the authors' original text. The translated version for processing in English (Table 6) can also be found in the list of tables.

Word2vec's Results: Non-significant

The results obtained from Word2vec algorithms lack robustness. There were several errors found in the orders that were identified automatically.

The analysis of textual corpora with word2vec algorithms did not provide significant results for three reasons. The first reason is the lack of a sufficient quantity of data. The second reason is that the posts were produced within communication strategies of publishing the results for COESO. This setting did not provide enough material where conflict situations of orders of worth could be identified. Therefore, our deliverable does not provide any further result about this algorithm. However, the list of keywords is useful as a preliminary exploration for future development that requires more data available. The third reason is that the model was simple for the accurate identification of conflicting orders of worth, as it was limited to matching keywords with the sentences of each speaker without providing any link to the whole argument. For instance, the table 4 below shows the identification of three orders "inspiration, opinion and civic" in the text extracted from the hypotheses' posts of Pilot 3.

Table 7. Three orders of worth with their corresponding keywords according to sentences from hypotheses' blog

Inspiration:	Matches: ['faire', 'corps', 'je', 'faire', 'faire', 'faire', 'faire']
	la journaliste , il s agit de faire l expérience de se « constituer une expertise au fil de la recherche , en partant d une base de connaissance la plus « neutre possible , pour pouvoir ensuite analyser son cheminement dans le cadre d une restitution sur le processus de notre recherche .

Opinion:	Matches: ['journaliste', 'orienter', 'grand public', 'journaliste', 'journaliste', 'journaliste', 'cause', 'journaliste', 'journaliste', 'journaliste', 'presse', 'journaliste']
	celles et ceux qui connaissent déjà notre projet , nous avons déjà dit que cette recherche est menée par deux profils différents : un chercheur en sciences politiques et sociales et une journaliste .
	, pour le projet , il était justement important qu'il y ait un regard suffisamment neuf pour que le chercheur puisse apprendre à orienter la restitution de son en se concentrant principalement sur les points qui peuvent intéresser le grand public et donc sortir de sa bulle d'expert , celle qu'il côtoie depuis des années et où il peut se retrouver n qu'avec des interlocuteurs « qui savent déjà de quoi il parle .
Civique	Matches: ['base', 'solidarité', 'justice', 'justice', 'loi', 'lutte', 'comité', 'comité', 'base', 'loi']
	la journaliste , il s'agit de faire l'expérience de se « constituer une expertise au fil de la recherche , en partant d'une base de connaissance la plus « neutre possible , pour pouvoir ensuite analyser son cheminement dans le cadre d'une restitution sur le processus de notre recherche .

There are two major issues we can observe. In the table there are few matches of words per order and they are repetitive. The keywords are not discriminant enough between orders. Therefore, the sentences are identified several times in different orders. The analysis of the second round of pilots did not provide significant results either.

4.2. Second technique: Sentence-Transformers and zero-shot classification

The second algorithms tested were based on a combination of sentence-transformers model and zero-shot classification. They were tested on pilot 7's meeting recordings that were transcribed and translated.

First, for Semantic Textual Similarity, we used the sentence-transformers model which will transform sentences into a dense vector space composed of 768 dimensions. The model used, called "stsb-mpnet-base-v2", is publicly available³. Using this model, we processed each sentence from the meeting and calculated cosine similarity with a provided list of keywords for each argument. This list of keywords corresponds to the dictionary previously built within our theoretical framework. The idea of using sentence-transformers model is, in comparison with

³ Source: <https://huggingface.co/sentence-transformers/stsb-mpnet-base-v2>

our previous attempt with word2vec approach, to take a sentence rather than each word as a unit of analysis to identify the idea conveyed in each sentence. This approach helps identify how closely related the content of each sentence is to the specified keywords.

In other words, this model helps gauge how similar sentences are in meaning. It is like a tool that reads each sentence from a meeting and checks how closely it aligns with a given list of important words related to each argument. The model helps to understand the context of discussions by measuring the similarity of sentences to key terms.

Second, a zero-shot classification was implemented in parallel to classify the sentiment in each sentence. This approach was undertaken as a response to another challenge encountered in the previous attempt ; identification of order of worth itself does not reveal a moment of conflict. Inspired by the model available at facebook/bart-large-mnli, we applied a zero-shot classification technique to label each sentence without any training dataset. This involves pre-classifying each sentence into one of three exclusive categories: 'agree,' 'disagree,' or 'neutral.' The goal was to help process a large quantity of email exchanges in an automatic way for pre-identifying moments of agreement or disagreement within the corpus for the researcher's post analysis. The model, trained on various natural language inference tasks, is employed to classify sentences without prior training on the specific task at hand. This implies that the data was directly put into the machine for recognition of orders of worths.

In other words, this model quickly sorts sentences into buckets based on their stance—whether they align, differ, or remain neutral. This way, we can pinpoint moments of consensus or divergence in the overall body of text.

Sentence-Transformers and zero-shot Results: Non-significant

Contrary to our expectation, the result obtained from the second approach did not significantly improve the identification of conflictual moments where a confrontation of different orders of worth is observed. Difference sentences identified by this model were still close to the ones identified by the model using Word2vec, and the identification still lacks robustness.

By examining the results, three possible reasons were raised to understand the difficulties in identifying a potential moment of conflict in the data.

Firstly, the output reveals that a disagreement is highly dissociated from a conflict in the analysed data. A close check of different moments of disagreement reveals that a negative sentence on a factual event is classified as a moment of

disagreement. For instance, an informal conversation where a member informs the rest of the team that another member is not attending the meeting can be classified as a disagreement even though no relevant order of worth can be identified.

Secondly, the scarcity of the target conflict moment in the data is a crucial factor. While developing a model capable of detecting a moment of conflict with different types of justifications, the existence of such a moment is not at all estimated. In order to come up with potential moments, the threshold of detection, such as a probability associated with the classification, was lowered ; this clearly induces a strong bias in the accuracy. As described concerning the email data, the presence of conflict, or any explicitly explained justification, might be simply missing or too 'rare' to be detected in the team discussion as well. The scarcity of the target category is a challenge when developing an algorithm which is supposed to find one. The original work of Boltanski and Thévenot detected conflicts of justification only within textbooks used as reference for each order of worth, and they mostly appeared as explicit critiques. Our requirements are more of a pragmatist kind, the analysis dealing with real life exchanges. This can partly explain the challenges we faced.

Thirdly, the unit of analysis should be adopted according to the type of data. In well written and structured text data such as official documents, each sentence or paragraph conveys different ideas. A sentence or a paragraph can therefore be considered as a meaningful unit of analysis to compare the similarity to detect orders of worth. However, in case of oral team discussion, the meaning of each sentence is determined in a larger context, where different speakers interact. The interaction plays a central role to convey different ideas, and an opposition or a conflict might not be explicitly understood in a single sentence. Contrary to structured text data, a clear unit of analysis is not identifiable ; there is no meaningful group of texts like a paragraph or a document. In this type of data, a categorization is necessarily based on partial information, which leads to a wrong classification or an omission of important information. The inaccuracy in the identification of order of worth might be partly related to this issue, namely that the model is not capable of finding a justification because of the sentence-based split of text data.

The script is available here:

[https://github.com/yuma-ando/COESO_Cooperation_Analytics/blob/main/Notebook/Zeroshot sentiment analysis with bert similarity Italian.ipynb](https://github.com/yuma-ando/COESO_Cooperation_Analytics/blob/main/Notebook/Zeroshot%20sentiment%20analysis%20with%20bert%20similarity%20Italian.ipynb)

5. Results Orders of Worth Pilot 7

Even though the original idea was to develop a semi-automated way of understanding the text data from an existing and pre-trained model, the lack of robustness and the presence of non coherent detection in the previous results suggested that the data should also be relevant for our development effort. We therefore decided to conduct an in-depth qualitative reading of the original text data by a specialist to further understand the underlying issues. This in-depth qualitative study was conducted by sociologist researcher Dominique Boulier in the original language of the team meeting, namely in Italian.

This section will firstly present the result of the qualitative analysis of the text by an expert, and secondly highlight the major challenges concerning how the manual and in-depth annotations of a corpus can be technically translated into a computational method. The analysis of the trends in discourses for each pilot was not possible as planned in task 5.3 given the limitations explained in this section. However, we found major lessons learned that we present for future research at the end of the section.

5.1. Manual identification of conflict situations and classification of orders of worth

As part of task 5.3 we were looking to identify local portraits of specific controversies and of their eventual resolution. After building a theoretical model, with a list of keywords and signals we were more specifically, identifying conflict situations between orders. Our first insight was to get into the material only when some conflicts were occurring or when the various orders of worth classified within the same data set and sequence of communication were significant. And then the manual analysis will take place, to interpret the situation and its motivations, its significance for the cooperation process.

Since we did not get any correct model to detect either the right orders of worth or the situations of conflicts, we decided to proceed with a qualitative approach. The expert would read the text extracted from the dataset of pilot 7 (only this one for tests purposes) and check the classification of orders of worth that can be done and at the same time detect where the situation was on the verge of conflict.

The classification of orders of worth worked pretty well manually, provided that the reviewer has a good knowledge of the topic and goal of the pilot. This is where the linguistic models are still lagging behind, since they do not have any semantic capacity and no relationship of any kind with the “real world” of actors and its

own ontology, its own dynamics of relationship between members and so on. The statistical work of classification can then be improved by this input from human expertise so that the vocabulary gets more precise and specific to the context.

Given the limitations of identifying conflicts, we were not able to find eventual resolutions of those conflicts for pilot projects. However, interesting lessons came from the detection of conflict or of moments of justification. Some conflicts may appear within the same order of worth (when a member criticises another member about his/her loyalty and refer to the superior requirements of trust in a group, or about the productivity and refer to the superior principle of collective efficiency and delivery constraints. Other conflicts are referring to different orders of worth but very often are not so conflictual because they are part of the challenges of assembling diverse people and fulfilling different types of expectations and requirements from various stakeholders. We may say that in citizen science, this is the most common situation when one tries to combine quality of scientific results and ethical principles of cooperation with citizens. But this is not considered as conflictual, since it is the everyday job of making compromises (this concept is a key feature of the original theoretical model and maybe we underestimated it during our investigation. In citizen science, one learning outcome might be to learn how to negotiate with these conflicting purposes and values without getting into a real conflict. We did not find any situation of radical conflict that could make a project split or fail (however this does exist in cooperation processes).

A final remark may help understand the problem of detection of conflict in cooperation processes. One clear indicator of a tension, a radical critique or a real conflict was the change in language formalism. To be clear, this is when members start using slang or bad language. It does not mean a lot about the type of conflict as such but when one stops over these shifts of styles of expressions, with a deeper investigation, it is quite easy to understand whether it is a conflict of judgments within one order of worth (rather easy to solve) or a conflict between different orders of worth which is much more difficult to handle since the perception of the situation differ radically.

5.2. Limits of Orders of Worth's technical implementation

An in-depth manual analysis of the data confirms the technical challenges in the implementation of orders of worth for identifying moments of conflicts in textual corpora. Consequently, a comparison of those conflicts with the cooperation analytics to detect patterns of online behaviour was not feasible as initially planned.

Firstly, as the example above showed, a sentence is not always the unit of analysis. A justification can be expressed only in a small part of a sentence, or sometimes over multiple sentences. This is due to the fact that the text data we are dealing with comes from an informal context of a team meeting. Members speak instantaneously in an interactive environment; the text data can only be understood in a context, where a meaning can be only fully understood in the interaction or other behaviours such as hesitation or repetition can play an important role. Another typical example of the technical limit is an implicit expression, when the sentence is not even terminated (e.g. “But it also means... You know what I mean.”). This type of sentence cannot be simply transformed into any vector, because the word is simply missing. A larger set of sentences should therefore be used as the unit of analysis, but no explicit rule can be established in oral conversations how to delimit the meaningful unit of analysis. This is even true as the qualitative study also showed in some cases that only a group of dozens of sentences revealed a meaningful understanding of different orders of worth at stake.

Secondly, the conflict itself is rarely manifest, and even avoided. While this was exactly the target moment to be detected, such a moment is scarce in the data. An algorithm of detection can only be efficiently developed and tested when the target category is clearly identified. Given the limited amount of data, we were unable to work on informative data on order of worth. If there were any citizen science project where a strong internal conflict hampered a successful achievement of the project, the implementation of the algorithm would have been more robust since it would have been more explicitly tested.

Finally, the translation between a human-based qualitative reading and the codification of the result is not straightforward. The different models that were tested are based on a specific task ; comparison of similarity or classification into specific categories. One possibility of improving the algorithm we tested was to fine-tune the existing model by feeding the qualitative study’s result as extra training data. However, in order to fine-tune the model, the output of the qualitative study should be mathematically expressed, for example by a binary coding with 1 indicating a presence of conflict and 0 if not. Given the limit related to a unit of analysis and the ambiguous nature of the real conversation (e.g. “not necessarily a conflict”), the qualitative study could not be used for fine-tuning. This discrepancy between the specific and explicit task of different computational methods and in-depth but more nuanced readings from an expert highlights another technical limit.

Conclusion: Lessons Learned

The development of task 5.3 allowed us to conduct a feasibility study of algorithms' development that is useful for future research. The approach was novel in terms of conceptualisation of sociological theories for the use of natural language processing techniques in citizen science. The dictionaries were compiled from Boltanski and Thévenot (who labelled them into each order of worth) and then we extended it with various NLP tests in order to process the whole corpus, from similarity detection. However, we faced data collection and technical limitations for going in-depth in the analysis of pilots' textual corpora as planned, namely identifying portraits of controversies for pilots was not possible. Consequently, we did not achieve going to a next phase for the analysis of pilots' conflicts and their eventual resolution. The comparison of their cooperation practices was only possible within the development of indicators in cooperation analytics (see deliverable 5.2) but not within the framing of the orders of worth.

At the theoretical level, the application of the theoretical framework of orders of worth to analyse text corpora from citizen science pilot projects was limited. Despite the linguistic subtlety of the approach, its technical implementation proved challenging, as well as its identification given the content of the exchanges between members in a team that were transmitted to WP5. Indeed, the expectation that cooperation would mobilise orders of worth did not unfold as anticipated. Calculating these orders of worth, while aimed at accountability for citizen science projects, did not emerge as the primary focus for pilots. There was a limited occurrence of moments of conflicts and a scarcity of explicit justifications, even with the extensive manual and qualitative analysis we performed on the data collected. Such scarcity hindered the adaptation of our theoretical framework and our desired depth of analysis. Unlike the theoretical framework based on orders of worth's conflicts, we observed a lack of generalisation and formality in the citizen science projects under study. A tendency to remain at a low level of justification made it difficult to identify negotiations between orders of worth. We observed in the pilots' exchanges mainly moments of finding compromises in project management or local arrangements; rather essential components of genuine cooperation and not formal conventions of scientific practices.

The complexity of citizen science and its accountability, comparable to classical science, is underscored by the need for explicit explanations in scientific procedures and publications. Challenges arose in explicitly validating the mobilised orders of industrial (i.e. where academia can be found), raising questions about the level of formality of citizen science at this point. Attempts to place citizen science within a conventional framework of orders of worth might suggest that there is a

predominant civic order in question over scientific rigour. As of today, there are myriad implicit processes in citizen science that are hindering the identification of distinct citizen science cooperation practices. The absence of formal identification procedures and the pilots' reliance on implicit approaches calls for a future effort to formalise citizen science, providing clearer criteria for justification in this domain.

At the computational level, for the purpose of implementing a theoretical framework in citizen science practices, various computational approaches and Natural Language Processing models were tested (keyword search, construction of semantic vectors and application of few-shot and zero-shot learning models). However, none of them provided robust results.

The first reason is that the detection of such communicative situations is too sophisticated for existing technologies and it cannot happen without human intervention. Therefore, the approach taken in WP5's project development was a semi-automated calculation performed after a first manual intervention from domain-experts identifying the conflict situations and classifying it within specific orders of worth to train models that could serve for further development. The starting point was two dictionaries of orders of worth, in English and in French. After the samples of potential examples of conflict situations were sent by the pilot projects, we processed the data according to the dictionaries manually built from the theory.

The second reason is that those models require a high volume of data that we were not able to collect. We faced limitations in the data collection process such as the variety of platforms where pilots run their activities, different data formats, and the lack of standardisation of processes which lead to receiving a lot of unstructured data.

The third reason was the variety of languages proper to pilot projects, this posed additional challenges for text processing. The techniques available for translation were not accurate enough. For instance, after analysing the translation of Italian text to English, the sense of the certain arguments was lost. In future research, a minimum of diversity should be ensured, and ideally focus on one language first for obtaining better results in natural language processing. At the current stage of technologies available, English is the most accurate language that can be analysed.

All these limitations must be considered as recommendations for next steps in the directions of understanding better the stakes of cooperation. This is specifically the benefit of deploying NLP techniques with the requirements of robustness that are rarely obtained in qualitative research.

References

- Boltanski, Luc, and Ève Chiapello. 2011. *Le nouvel esprit du capitalisme*. Paris: Gallimard.
- Boltanski, Luc, and Laurent Thévenot. 1999. "The Sociology of Critical Capacity." *European Journal of Social Theory* 2(3): 359–77.
- Boltanski, Luc, and Laurent Thévenot. 2006. *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Annex

1. Data Collection Guidelines for COESO's Pilot Projects



COESO
connecting research and society

Data Collection Guidelines for COESO's Pilot Projects

WP5- Cooperation Analytics

By Dominique Boullier, Jessica Pidoux

Guidelines:

1. Meetings

- Minutes for each meeting should be written down
- The minutes' meetings should use COESO template
- Shared them with us in a document, see section 6 *Documents* for more details

2. Calendars

- A calendar app should be used for invitations to meetings, in the invitation include the **topic** meeting
- For calendar invitations, in the list of invited users, add in C.C your corresponding generic email for the cooperation analytics:
 - coesopilot6@sciencespo.fr
 - coesopilot7@sciencespo.fr
 - coesopilot8@sciencespo.fr
 - coesopilot9@sciencespo.fr
 - coesopilot10@sciencespo.fr
- For in-depth analysis you have to export the calendar data in the **parameters** section of your calendar.

3. Mails

- When sending COESO-related mails, add in C.C your corresponding generic email for the cooperation analytics:
 - coesopilot6@sciencespo.fr
 - coesopilot7@sciencespo.fr
 - coesopilot8@sciencespo.fr
 - coesopilot9@sciencespo.fr
 - coesopilot10@sciencespo.fr
- Mail subjects should include the word [COESO] for data analysis purposes

4. Discussions

- Mattermost is fully integrated in VERA. When a project is created in VERA a private dedicated channel is activated in the OPERAS Mattermost instance, accessible at the URL <https://messaging.operas-eu.org/>. All members of the project team are automatically added to this channel, which is therefore ready to be used for managing the communication in your pilot.
- It is strongly suggested that you use it for your pilots' internal communication. We are exploring the possibility of providing pilots an analysis of their conversations online if you use this platform. Please notify WP5 if you are interested in such analysis.

5. Profile Page

- Every pilot should complete a profile page that will help obtaining basic project information for the cooperation analytics' tests
- A profile page for your cooperation analytics should be filled in by every member of your team
- When new project members join, they should complete the profile form above
- If any current member is leaving the project, please notify WP5

6. Documents

- **Share the minutes of your meetings and any other document relevant for text analysis (including recorded meetings)** that concerns your working practices. We provide two options for sharing your documents with WP5:
 - Use SharedDocs for storage purposes if you don't have, or if you want, a dedicated storage for cooperation analytics.
 - If files are already stored in your personal or professional cloud, drive, or local storage devices, provide access to the relevant folders for analytical purposes to your dedicated email (see section 3 *Mails*).

How to use ShareDocs:

- 1) Designate a person responsible for data storage and to copy the documents related to your pilot project in ShareDocs
- 2) This person has to create first a Huma-Num account <https://humanid.huma-num.fr/>

- 3) Once the account is created in Huma-Num, login and request a ShareDocs account, **specify in your request that you want to be part of the shared folder**



Plateforme de
stockage et de
partage de fichiers
(Web et clients
WebDAV)

- 4) Inform WP5 this person's name and email address
- 5) After your ShareDocs' account is approved you can now start using the shared repository with COESO's WP5 and add your documents for analytical purposes

Why cooperation analytics can be helpful for all Pilot projects

When doing citizen science, we are aware that the process of knowledge production becomes complex. Citizen and professional scholars have to cooperate to deliver new insights that traditional scientific practices would have missed or disqualified. They have to conciliate different standpoints on a specific issue, as well as their multiple protocols to validate knowledge. That is why all stakeholders of a pilot project must take into account the quality of the cooperation process. Our definition of quality does not refer to criteria defined in a self-evaluation survey at the end of the project. This will not help stakeholders to gain a better understanding of their research process. The COESO project advocates precise and useful feedback for pilots within the VERA platform as to learn about their on-going cooperation practices; that is to observe citizen science cooperation "in the making". The feedback will be given in the form of a dashboard displaying cooperation analytics.

For developing the cooperation analytics, we would like to test the relevance of 30 indicators for measuring cooperation that we designed from the literature and adapted to the citizen science practices. Pilots are the only ones in the position to validate our prototype. But since we do not want to put any extra cognitive burden on the stakeholders, we managed to design indicators that are automatically calculated from the traces of everyday online interaction that we extract from VERA. Only the project's members will choose how to use and make sense of the cooperation analytics. The cooperation models we developed are not normative, and they do not include rankings, or scoring purposes of any kind. Our ultimate

goal is to offer project's members scaffolds for reflexivity about the evolution of citizen science cooperation during the research.

In order to collect reliable and sufficient amount of data, pilots are requested to adopt some platforms and protocols that will help us to provide you useful insights about your project. These recommendations allow us to deploy our calculations in VERA with algorithms that we tested already with the first group of pilots. They might be part of your basic COESO agreement, or what we need in WP5 for the cooperation analytics to be computed. The recommendations can also be useful for your own coordination and formalisation of the scientific process.

These recommendations will be discussed on a one-to-one basis with each pilot to understand their constraints, and if necessary to set up a guidance process or to find an arrangement with the existing uses. The data we collect for computing the cooperation analytics follows our Data Management Plan that complies with the General Data Protection Regulation (GDPR).

Document's versioning

Date	Update	Version
26.06.2022	Original	v1
10.08.2022	Shorten text for better readability	v2
16.01.2023	Dedicated mails to pilots, new section calendars and more options all sections	v3

List of tables

Table 1: Six orders of worth

	Inspired	Domestic	Civic	Opinion	Market	Industrial
Mode of evaluation (worth)	Grace, non conformity, creativeness	Esteem, reputation	Collective interest	Renown	Price	Productivity, efficiency
Format of relevant information	Emotional	Oral, exemplary, anecdotal	Formal, official	Semiotic	Monetary	Measurable, criteria, statistics
Elementary relation	Passion	Trust	Solidarity	Recognition	Exchange	Functional link
Human qualification	Creativity	Authority	Equality	Celebrity	Desire, purchasing power	Professional competency, expertise

Source: L. Boltanski and L. Thévenot, 1999, "The sociology of critical capacity", *European Journal of Social Theory*, vol.2, n°3, special issue "Contemporary French Social Theory", (August).

Table 2. Data collected per pilot based on the different sources exploited in the guidelines

Pilot No.	Meeting minutes	Meeting recorded	Emails	Profile page	Calendar	Hypotheses
Pilot 6	no	no	no	yes	no	yes
Pilot 7	yes	yes	yes	yes	no	yes
Pilot 8	yes	no	yes	yes	no	yes
Pilot 9	yes	no	yes	yes	no	yes
Pilot 10	yes	no	yes	yes	no	yes

Table 3. Mail dataset

Pilot No.	Qty.	Period
6	0	N/A
7	47	Aug.2022-Feb.2023
8	35	Jul.2022-Nov.2022

9	74	Jul.2022-Nov.2022
10	62	Oct.2022-Jan.2023
COESO WP5 members	196	Mar.-Jul. 2022

Table 4. Meetings dataset

Pilot No.	Qty.	Period
6	0	N/A
7	8 mp4 videos, 4 Word documents, 17 PDF documents	Sept. 2022-Apr. 2023
8	7 Word documents	Jun.-Nov. 2022
9	10 Word documents	Jul.-Nov.-2022
10	0	N/A

Table 5. Principles of seven orders of worth in French

1/ inspiration	2/ domestique	3/ opinion	4/ civique	5/ marchand	6/ industriel	7/ projet
inspiration	tradition	la réalité de l'opinion	la prééminence des collectifs	concurrence	l'efficacité	projets
indicible et éthéré	hiérarchie	les autres	collectif	rivalité	performance	extension du réseau
bizarre	génération	grand public	tous	compétition	avenir	prolifération des liens
insolite	la supériorité hiérarchique	la célébrité	volonté générale	désirable	fonctionnel	connecter
merveilleux	bienveillant	réputé	réglementaires et représentatifs	vendable	performant	communiquer

indicible	bien élevé	reconnu	unitaire	valeur de	fiable	se coordonner
inquiétant	aisé	visible	légal	millionnaire	opérationnel	s'ajuster aux autres
passionnant	distingué	succès	réglementaire	gagneur	energie	faire confiance
spontané	discret	avoir du succès	officiel	détesté	travail	engagé
émotionnel	réservé	se distinguer	représentatif	inadapté	professionnels	engageant
l'inquiétude de la création	digne de confiance	persuasif	autorisé	aléatoire	experts	mobile
amour	franc	accrocheur	titulaire	l'intérêt	spécialistes	enthousiaste
passion	fidèle	le désir d'être reconnu	libre	désir	responsable	impliqué

créer	l'aisance de l'habitude	amour-propre	l'aspiration aux droits civiques	égoïsme	opérateur	flexible
les illuminés	pli	considération	droits civiques	concurrents	outils	adaptable
esprit	bon sens	désir de	aspirations politiques	homme d'affaires	ressources	évolutif
fée	naturel	les vedettes et les supporters	participation	vendeur	méthode	polyvalent
ombre	caractère	une personnalité	collectivités publiques	client	tâche	employable
monstre	les supérieurs et les inférieurs	leader d'opinion	parti	acheteur	espace	autonome

je	père	porte-parole	fédération	travailleur indépendant	environnement	non prescrit
enfant ⁴	roi	relais	section	richesse	axe	à l'écoute
femme	ascendants	journaliste	bureau	objets de luxe	direction	tolérant
fou	parents	attaché de presse	comité	opportunisme	critère	médiateur
artiste	famille	des noms dans les médias	élu	liberté	définition	chef de projet
le rêve éveillé	personne	marque	représentant	ouverture	liste	coach
corps	patron	message	délégué	attention aux autres	graphique	expert
rêve	chef	emetteur	secrétaire	dynamique	schéma	client

⁴ Enfant (Child, in english) appears two times: in the inspiration order where it is related to spontaneity and creativity, in the second order of domestic where it is related to generations

inconscient	enfant	récepteur	adhérent	distance émotionnelle	calendrier	fournisseur
drogue	femme	campagne	droit	prendre du recul	plan	innovateur
l'évasion hors des habitudes	étranger	relations publiques	législation	posséder	objectif	instruments de connexion
remettre en question	célibataire	presse	décret	intéresser	quantité	technologies nouvelles
risque	je	interview	ordonnance	acheter	variable	relations informelles
détour	moi	communiqué	mesure	produire	série	relations de confiance
la valeur universelle de la singularité	chien	support	tribunaux	vendre	moyenne	partenariat
indépendant	chat	brochure	formalité	être en affaires avec	probabilité	accords

génie	autres	mailing	procédur e	négociier		alliances
l'alchimie des rencontres imprévues	visiteur	badge	protocole d'accord	tirer parti	facteur	sous- traitance
découvrir	entourage	audio-v isuel	dérogatio n	monnayer	cause	réseaux d'entreprise
rechercher	voisins	ambian ce	capacité électoral e	payer	progrès	entreprises en réseau
imaginer	tiers	décor	code	rivaliser	investissem ent	maillé
rêver	les règles du savoir-vivr e	le renonc ement au secret	critère	marché	sympathie	boucle neurone
exploser	bonnes manières	révéler	circonscri ption	affaire	détacheme nt	synapse

faire	bienséance	être reconnu	liste électorale	affaire réglée	maitrise	projets
la réalité de l'imaginaire	rang	s'identifier	programme	dans le sac	fonctionner	interchangeable
imaginaire	titre	identification	orientation	marché conclu	mettre en œuvre	inadaptable
inconscient	demeure	force	déclaration	prix	liaisons d'engrenage	n'inspire pas confiance,
usages	présentation	persuasion	affiche	valeur justifiée	être fonction de	autoritaire
anticonvencances	signature	influence	brochure	valeur raisonnable	rouage	rigide
aventure	faire-part	convaincre	bulletin	valeur vraie	avoir besoin	intolérant
quête	cadeaux	sensibiliser	tract	argent	conditionner	immobile

voyage mental	fleurs	accrocher	slogan	bénéfice	relation nécessaire	local
expérience vécue	le rejet de l'égoïsme	séduire	siège	résultat	intégrer	enraciné
illumination	serviabilité	percer	permanence	rétribution	organiser	attaché
intuition	devoir (et d'elle)	capter	local		contrôler	statut
faillir apparaître	harmonie	lancer	sigle		stabiliser	sécurité
vertige	respect et responsabilité	émettre	carte		prévoir	fermeture du réseau
se dépasser	autorité	circuler	solidarité		implanter	corruption
chef d'œuvre	subordination	propager	dépasser		adapter	copinage
planer	respectabilité	débattre	renoncer		détecter	privilèges

aura	honneur	prendre parole	lutte		analyser	mafias
fantasme	honte	promouvoir	les rapports de délégation		prendre en compte	redistribution des connexions
la certitude de l'intuition	le commerce des gens bien élevés	orienter	adhésion		déterminer	mettre en contact
symbole	reproduire	amplifier	représentation		mettre en évidence	redistribuer l'information
signes	enfanter	parler de	délégation		mesurer	insérer dans des réseaux
analogie	éduquer	citer	traduire les aspirations		formaliser	donner de l'employabilité

Images	inviter	l'image dans le public	unifier		standardisation	adaptabilité
mythes	donner	audience	mobiliser		optimiser	légèreté
	recevoir	cible	rassembler		résoudre	tolérance
	rendre	positionnement	exclure		traiter	location
	recommander	la présentation de l'événement	adhérer		organisation	flexibilité
	remercier	manifestation	se rallier		système	la fin d'un projet
	respecter	conférence de presse	lancer un appel		test	le début d'un autre projet

	l'âme du foyer	imagination	débattre		lancement	être appelé à participer
	famille	porte ouverte	prendre la parole		mise en route	insérer
	milieu	l'évidence du succès	informer		mise en œuvre	faire participer
	principe	connu	codifier		réalisation	parler de
	fête	répercussion	légaliser		effectif	éviter
	naissance	réduire à de	habiliter		correct	tenir à l'écart
	décès	cause	saisir les tribunaux		en ordre de marche	ignorer
	mariage	se rallier à	la république démocratique		fonctionnant	rejeter

	mondanité	prise de conscience	république		mesure	exclure
	conversation	media	état			besoin de se lier
	distinction		démocratie			réseau
	nomination		base			
	apprécier		électorat			
	féliciter		institutions représentatives			
	remontrances		parlement			
	rapporter		assemblée			

	l'anecdote exemplaire		congrès			
	donner en exemple		conseil			
	préjugé		réunion			
			session			
			mouvement			
			manifest er la présence			
			litige			
			recours			
			justice			
			le verdict du scrutin			
			retentiss ement			

			vote			
			election			
			consultati on			
			mobilisati on			
			loi			
			règles juridiques			
			statuts			

Table 6. English Dictionary with seven orders of Worth and their corresponding keywords and signals of expressions per order

1/ inspired	2/ domestic	3/ opinion	4/ civic	5/ merchant	6/ industrial	7/ project
inspiration	generation	the reality of opinion	the pre-eminen	competition	efficiency	projects

			ce of collectives			
unspeakable and ethereal	hierarchy	others	collective	rivalry	performance	network expansion
weird	tradition	General public	all	competition	coming	proliferation of links
unusual	hierarchical superiority	Fame	general will	desirable	functional	connect
marvellous	caring	famous	regulatory and representative	salable	efficient	communicate
unspeakable	well-mannered	recognized	unitary	value	reliable	coordinate
disturbing	easy	visible	legal	millionaire	operational	adjust to others
exciting	distinct	hit	regulatory	winner	energy	trust
spontaneous	discreet	be successful	official	hated	work	engaged
emotional	reserve	stand out	representative	unsuitable	professionals	engaging
concern for creation	trustworthy	persuasive	allowed	random	experts	mobile

love	franc	catchy	incumbent	interest	specialists	enthusiastic
passion	faithful	the desire to be recognized	free	desire	responsible	implied
create	ease of habit	self-esteem	the aspiration to civil rights	selfishness	operator	flexible
the enlightened	fold	consideration	civil rights	competitors	tools	adaptable
mind	common sense	desire to	political aspirations	business man	resources	scalable
fairy	natural	stars and supporters	participation	seller	method	versatile
shadow	character	a personality	public authorities	customer	stain	employable
freak	the superiors and the inferiors	opinion leader	left	Buyer	space	autonomous
l	father	spokesperson	federation	self employed	environment	not prescribed
child	king	relay	section	richness	axis	listening

women	ascendant s	journalist	desk	luxury items	direction	tolerant
mad	parents	Press officer	committee	opportunism	criteria	mediator
artist	family	names in the media	elected	freedom	definition	Project Manager
the waking dream	person	brand	representin g	opening	list	coach
body	boss	message	delegate	pay attention to others	chart	expert
dream	chief	transmitter	secretary	dynamic	plan	customer
unconsciou s	child	receiver	member	emotional distance	calendar	supplier
dope	women	campaign	right	take a step back	plan	innovator
escape from habits	foreign	public relations	legislation	possess	objective	connection tools
question	Single	press	decree	to interest	quantity	new technologies
risk	l	interview	arrangemen t	buy	variable	informal relationships
detour	Me	communica ted	measure	produce	series	trust relationships

the universal value of singularity	dog	support	courts	sell	mean	partnership
independent	cat	brochure	formality	to be in business with	probability	chords
genius	others	mailing	procedure	negotiate	nun	wedding rings
the alchemy of unexpected encounters	visitor	badge	protocol of agreement	benefit	factor	outsourcing
discover	surroundings	audio-visual	exemption	monetize	cause	business networks
to research	neighbours	atmosphere	electoral capacity	pay	progress	network companies
imagine	third	decor	coded	compete	investment	mesh
dream	the rules of good manners	reconcentration to secrecy	criteria	market	sympathy	neuron loop
explode	good manners	reveal	riding	affair	detachment	synapse
TO DO	propriety	to be recognized	electoral list	case settled	mastery	projects

the reality of the imaginary	rank	identify	program	in the bag	function	unchangeable
imaginary	title	identification	orientation	bargain	implement	unsuitable
unconscious	remains	strength	statement	price	gear linkages	does not inspire confidence,
uses	presentation	persuasion	attach	justified value	be a function of	authoritarian
proprieties	signature	influence	brochure	reasonable value	cog	rigid
adventure	announcement	convince	newsletter	true value	need	intolerant
quest	gifts	to raise awareness	leaflet	money	to condition	motionless
mental journey	flowers	hang	slogan	profit	necessary relationship	local
lived experience	rejection of selfishness	seduce	seat	result	to integrate	rooted
illumination	helpfulness	drill	permanence	retribution	arrange	attached

intuition	duty (and she)	capture	local		control	Status
almost appear	harmony	throw	acronym		stabilise	security
vertigo	respect and responsibility	issue	map		to expect	network shutdown
surpass oneself	authority	circulate	solidarity		implant	corruption
masterpiece	subordination	propagate	exceed		adapted	cronyism
hover	respectability	discuss	to renouncer		detect	privileges
will have	honour	speak up	struggle		analyse	mafias
fantasy	shame	promote	delegation reports		take into account	redistribution of connections
the certainty of intuition	the trade of well-mannered people	orient	membership		determine	get in touch

symbol	reproduce	amplify	representati on		highlight	redistribute information
signs	give birth	talk about	delegation		measure	insert into networks
analogy	educate	to quote	translate aspirations		formalise	provide employability
images	invite	public image	unify		standardisati on	adaptability
myths	give	audience	mobilise		to optimise	lightness
	receive	target	gather		solve	tolerance
	render	positioning	exclude		to treat	lease
	recommen d	the presentatio n of the event	join		organisation	flexibility
	thank	expression	rally		system	the end of a project
	respect	press conference	make a call		test	the start of another project
	the soul of the home	imagination	discuss		launch	to be called upon to participate

	family	open door	speak		getting started	insert
	environment	evidence of success	inform		Implementation	involve
	principle	known	codify		achievement	talk about
	party	repercussion	legalise		effective	avoid
	birth	reduce to	empower		correct	keep away
	death	cause	go to court		in working order	ignore
	marriage	rally to	the democratic republic		functioning	dismiss
	worldlines	awareness	Republic		measure	exclude
	conversation		state			need to bond
	distinction		democracy			network
	nomination		base			
	enjoy		electorate			
	congratulate		representative institutions			

	admonitions		parliament			
	report		assembly			
	the exemplary anecdote		Congress			
	give an example		advice			
	prejudice		meeting			
			session			
			movement			
			manifest the presence			
			dispute			
			appeal			
			justice			
			the verdict of the ballot			
			rumour			
			noise			
			fashion			
			coast			

			resounding			
			vote			
			election			
			consultation			
			mobilisation			
			measure audience			
			law			
			legal rules			
			statutes			

Table 7. Three orders of worth with their corresponding keywords according to sentences from hypotheses' blog

Inspiration:	Matches: ['faire', 'corps', 'je', 'faire', 'faire', 'faire', 'faire']
	la journaliste , il s agit de faire l expérience de se « constituer une expertise au fil de la recherche , en partant d une base de connaissance la plus « neutre possible , pour pouvoir ensuite analyser son cheminement dans le cadre d une restitution sur le processus de notre recherche .
Opinion:	Matches: ['journaliste', 'orienter', 'grand public', 'journaliste', 'journaliste',

	'journaliste', 'cause', 'journaliste', 'journaliste', 'journaliste', 'presse', 'journaliste']
	celles et ceux qui connaissent déjà notre projet , nous avons déjà dit que cette recherche est menée par deux profils différents : un chercheur en sciences politiques et sociales et une journaliste .
	, pour le projet , il était justement important qu'il y ait un regard suffisamment neuf pour que le chercheur puisse apprendre à orienter la restitution de son expertise en se concentrant principalement sur les points qui peuvent intéresser le grand public et donc sortir de sa bulle d'expert , celle qu'il côtoie depuis des années et où il peut se retrouver n'importe quand avec des interlocuteurs « qui savent déjà de quoi il parle .
Civique	Matches: ['base', 'solidarité', 'justice', 'justice', 'loi', 'lutte', 'comité', 'comité', 'base', 'loi']
	la journaliste , il s'agit de faire l'expérience de se « constituer une expertise au fil de la recherche , en partant d'une base de connaissance la plus « neutre possible , pour pouvoir ensuite analyser son cheminement dans le cadre d'une restitution sur le processus de notre recherche .