

March 2022

HyEnA: A Hybrid Method for Extracting Arguments from Opinions

Supplementary material

Michiel VAN DER MEER ^{a,1}, Enrico LISCIO ^b, Catholijn M. JONKER ^{a,b},
Aske PLAAT ^a, Piek VOSSSEN ^c and Pradeep K. MURUKANNAIAH ^b

^a*Leiden Institute for Advanced Computer Science (LIACS), Leiden University*

^b*Interactive Intelligence (II), TU Delft*

^c*Computational Lexicology and Terminology Lab (CLTL), VU*

A. Experiment Protocol & Description

In order to reproduce the experiments performed in this research, we provide a complete overview of the guidelines, preliminaries, data and technical artifacts created [1]. This overview contains additional information about how the experiments were conducted. The texts presented to the annotators, such as the informed consent, the annotation introduction and instructions are provided in the supplementary material as well. In addition, we provide details on the average run times per experiment, as well as any other auxiliary details here.

A.1. Preliminaries

Before starting the experiments, annotators were required to familiarize themselves with the annotation procedure and web interface. Upon entering the web platform, they were provided with an informed consent form and the instructions for their task. The instructions consist of short introduction to the context of the task, followed by detailed instructions about the components they would be annotating (opinions, arguments, topics, etc.). In addition, they were provided example annotations, both in writing and by means of a video.

After having seen all these, annotators were asked to fill in a short exercise annotation. This exercise consisted of 3 or 4 items, applicable to a hypothetical policy option, each with a predefined correct answer. Annotators were required to get the answers correct, but had unlimited tries to perform the exercise. Completing the exercise enabled the actual annotation task, which in all cases was upper-bounded by a fixed number of items. Annotators were paid £7,50 per hour which is considered an ethical monetary reward on Prolific.

¹Corresponding Author: Michiel van der Meer, Leiden Institute for Advanced Computer Science, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands; E-mail: m.t.van.der.meer@liacs.leidenuniv.nl.

Although the opinion corpora contain comments on Dutch policy, the annotators were not restricted to certain (geographical) demographics. 88% of the annotators resided in continental Europe at the time of annotation, with the next 9% residing in Middle and North America. The average age of annotators was 28 ($SD = 7.7$). For 71% of annotators, data on student status was available, and around half (36 p.p.) indicated currently being a student.

A.2. Phase 1: Argument Annotation

This first phase of HyEnA consists of three stages. We provide some additional details per stage. For the interpretation of the results, we refer to the original paper.

Argument Annotation Five annotators were given one hour to explore 51 opinions from the corpus for a single option. On average, they took 44, 31, and 43 minutes respectively for the options of YOUNG, IMMUNE and REOPEN.

Topic Generation Two experts worked to generate a short list of topics from the 15 most frequent BERTopic generated topics, with the short list containing only coherent and unique topics. Two experts worked for 23 minutes on average to rate all topics across all three options.

Topic Assignment In the topic assignment, each argument from the **argument annotation** stage had to be provided with a manual topic assignment. Topics are assigned by five overlapping annotators. For YOUNG, IMMUNE and REOPEN, they took 26, 30, and 33 minutes respectively on average.

A.3. Phase 2: Argument Consolidation

The arguments were consolidated by 99, 57, and 87 annotators for the options of YOUNG, IMMUNE and REOPEN respectively. The median completion time was 20, 20 and 18 minutes. In the Multi Path algorithm in use by POWER multiple annotators are able to work in parallel, supported by our annotation platform.

A.4. Comparison to Automated Baseline

Lastly, in the comparison between HyEnA and ArgKP, annotators rated a fixed number of opinions and arguments. For the option YOUNG, 28 annotators took 23 minutes on average. For both IMMUNE and REOPEN, both options saw 21 annotators, which took 25 and 23 minutes on average respectively. In this task, the annotators were asked to assess the match between arguments and opinions, where *matching* is defined as “an argument capturing the gist of the opinion, or directly supports a point made in the opinion.”

A.5. Annotation platform

To run the HyEnA experiments and employ the workers from Prolific (www.prolific.co), we created our own web platform that supports all phases in HyEnA. The platform allows annotators to work in parallel, and is equipped with control mechanisms for conducting the experiments. Where possible, computations are performed offline, which is possible for all phases with the exception of the Parallel Pairwise Annotation method, POWER. For this phase, we precomputed the dependency graph G , and extracted the

disjoint paths containing the pairs to be annotated. Following the annotator’s decisions, we then make automated judgements over sections of these paths. We add screenshots of the pages as presented to the annotators in the `screenshots/` directory.

The ArgKP baseline was run using two RTX 3090 Ti GPUs, which took around 30 hours per opinion corpus. For HyEnA, the opinion corpus was transformed into embeddings using the same device within 4 hours. Training the BERTopic models took less than an hour. All web-based experiments were hosted on a single server with 16GB RAM, without access to a GPU.

B. Method Details

B.1. Opinion Corpus

For an overview of the options, see Table 1. Opinions were entered by Dutch citizens in April 2020 following a Participatory Value Estimation (PVE) study. We manually split the data into separate corpora of opinions related to each of the options. Since some opinions entered in the original questionnaire were applicable to multiple options, we copy the opinion for all relevant options. We provide the full dataset of opinions, as well as the annotations performed by the annotators in HyEnA.

Table 1. Statistics for the three policy proposals (options) in the COVID-19 corpus.

Policy option	Size	Pro/Con ratio
YOUNG people do not need to maintain 1.5 meter distance among each others	13400	0.66/0.34
All restrictions are lifted for persons who are IMMUNE	10567	0.17/0.83
REOPEN hospitality and entertainment industry	12814	0.55/0.45

B.2. Parallel Pairwise Annotation Algorithm

To accommodate annotators performing asynchronous annotation, we take an incremental procedure for pairwise annotation. As soon as a pair has seen three annotations, the automatic labeling procedure is run, and the next pair to be annotated in the same path is opened up for annotation. When all pairs are (either manually or automatically) labeled, the algorithm is complete. See Algorithm 1 for computational description of the parallel pairwise annotation algorithm [2]. Since the paths are annotated through a binary traversal method, we can also obtain an upper bound of number of annotations required, which is the number of paths $|P|$ multiplied by the maximum number of annotations required for the longest path g , $P \times \lceil \log_2(|g|) \rceil$.

B.3. Hyperparameters

B.3.1. HyEnA

An overview of hyperparameters for HyEnA is given in Table 3.

Algorithm 1: Parallel Pairwise annotation

Input: Dependency graph $G = \{V, E\}$
Output: Labeled vertices V

```

1  $B = \text{create bipartite graph}(G)$ 
2  $Y = \text{find maximal matching}(B)$ 
3  $P = \text{find disjoint paths}(Y)$ 
4 while  $\neg \text{fully labeled}(G)$  do
5   for  $p \in P$  do
6      $v = \text{find middle}(p)$ 
7      $\text{label vertex}(v)$  ;  $\triangleright N$  humans
8   end
9    $\text{automatically label paths}(P, \text{label})$ 
10 end

```

B.3.2. ArgKP

Table 4 shows the hyperparameters for the ArgKP baseline. The hyperparameters for the ArgKP baseline were picked such that they are balanced between the ones used for the Argument dataset [3], but also would increase (up to $\sim 10\%$) the ratio of comments picked as key point candidates. While this is lower than the recommended 20%, we avoided relaxing the heuristic hyperparameters to prevent picking overly specific arguments as candidates. In Figure 1, we show the ratio of number of candidates extracted out of all opinions depending on the hyperparameters.

Running ArgKP does not come cheap. The number of comparisons required to be made (forward passes through the matching model) is $\mathcal{O}(NM)$ where N is the number of candidates and M the number of opinions. Table 2 shows the number of comparisons made by the model in use in our experiments.

Table 2. Quantative descriptive information for running ArgKP.

Option	Stance	# Opinions	# Candidates	# Comparisons
YOUNG	pro	8804	1307	12M
YOUNG	con	4596	463	2M
IMMUNE	pro	1760	369	649K
IMMUNE	con	8807	657	6M
REOPEN	pro	7027	690	5M
REOPEN	con	5787	457	3M

C. Ethical Considerations

Our paper develops and evaluates a hybrid (human and AI) approach to extracting key arguments from an opinion corpus. The intended use case for our method is synthesizing key arguments that are grounded in opinionated policy-related comments, by using a pool of annotators. We identify two main aspects of risk in our method.

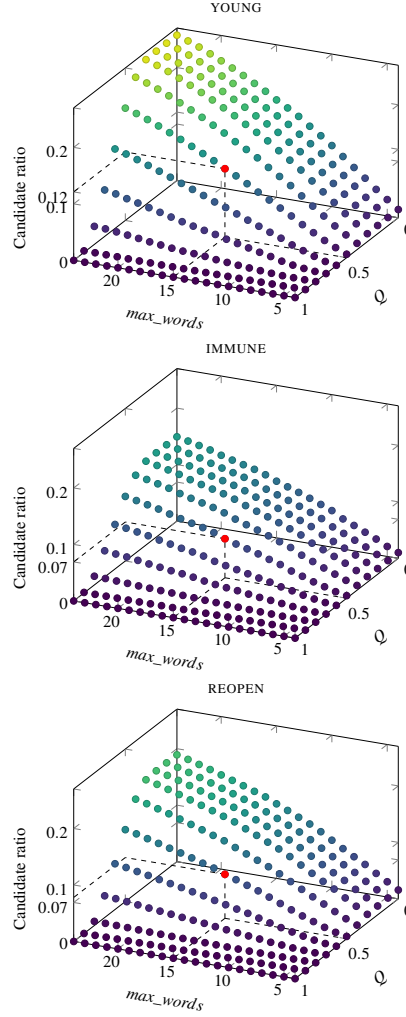


Figure 1. Hyperparameter sweep for ArgKP (max_words and Q) and its impact on the ratio of candidates picked. The indicated red dot shows the chosen parameter settings.

First, we aim to mitigate the effect of individual biases by grounding the key arguments in general public user opinions. However, the key argument extraction is ultimately performed by individual annotators. We address the influence of subjectivity and noise by combining multiple annotators in the consolidation phase. Further, as our method is transparent, the complete annotation process (from opinions to consolidated key arguments) is traceable. One could implement additional checks on annotator behavior as a bias-mitigating factor, which is a significant research challenge on its own.

Second, the diversity of the opinion embeddings is contingent on the representational quality of the S-BERT model. Underlying biases in its representation may influence the opinions sampled. However, we use FFT to actively sample diverse opinions, which can reduce the impact of inaccurate embeddings.

Table 3. Hyperparameters used by HyEnA.

Parameter	Option	Value	Description
M_{SBERT}	all	paraphrase-MiniLM-L6-v2	Model used to transform opinions and arguments into a numerical representation.
$\mathcal{T}$	all	paraphrase-MiniLM-L6-v2	Model in use by BERTopic.
f	all	5	Number of farthest opinions to sample using FFT.
clustering method	YOUNG	louvain	Clustering method used to extract argument clusters per option.
	IMMUNE	louvain	
	REOPEN	spectral	
r	YOUNG	0.449	Resolution parameter for Louvain clustering.
r	IMMUNE	0.449	Resolution parameter for Louvain clustering.
k	REOPEN	18	Number of desired clusters for spectral clustering.

Table 4. Hyperparameters for the ArgKP baseline used in the comparison against HyEnA. We also show the originally proposed values [3].

Parameter	Option	Value	Baseline Values	Description
min_words	all	1	1	Minimum number of words in an opinion to be considered a key point candidate.
max_words	all	15	10, 12	Maximum number of words in an opinion to be considered a key point candidate.
Q	all	0.5	0.4, 0.5, 0.7	Minimum argument quality according to a model trained on the ArgQ dataset [4].
θ	all	0.9	0.856, 0.999	Threshold value for match scores for (1) assigning opinions to key point candidates and (2) merging similar key point candidates.

D. Detailed Results

D.1. Unclear Translation Actions

In the argument annotation phase of HyEnA, when extracting arguments from opinions, annotators had the option to skip the opinion if they could not extract any argument from the opinion. Since opinions were automatically translated by the Azure translation service, we also made it optional to indicate that the reason for skipping the argument was because of an unclear translation. Out of 51 actions, annotators indicated mistranslations in 6, 7 and 2 opinions on average for YOUNG, IMMUNE and REOPEN respectively. This shows that the machine translation caused only some noise, and the majority of the skipped opinions were skipped because of different reasons (e.g. no argument was present in them).

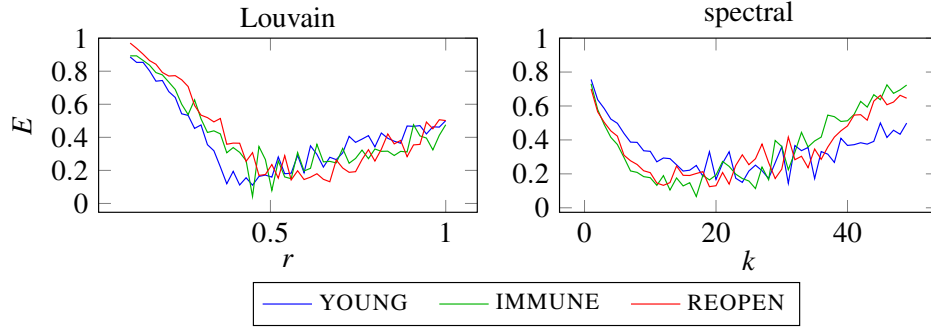


Figure 2. Parameter tuning in key argument clustering.

D.2. Clustering Arguments

D.2.1. Optimizing for E

Figure 2 show the optimal parameter setting for the clustering methods over each corpus. We also present an alternative visualization, now separated in Figure 3. The lowest observed score is indicated with the red line, obtained by the method in bold.

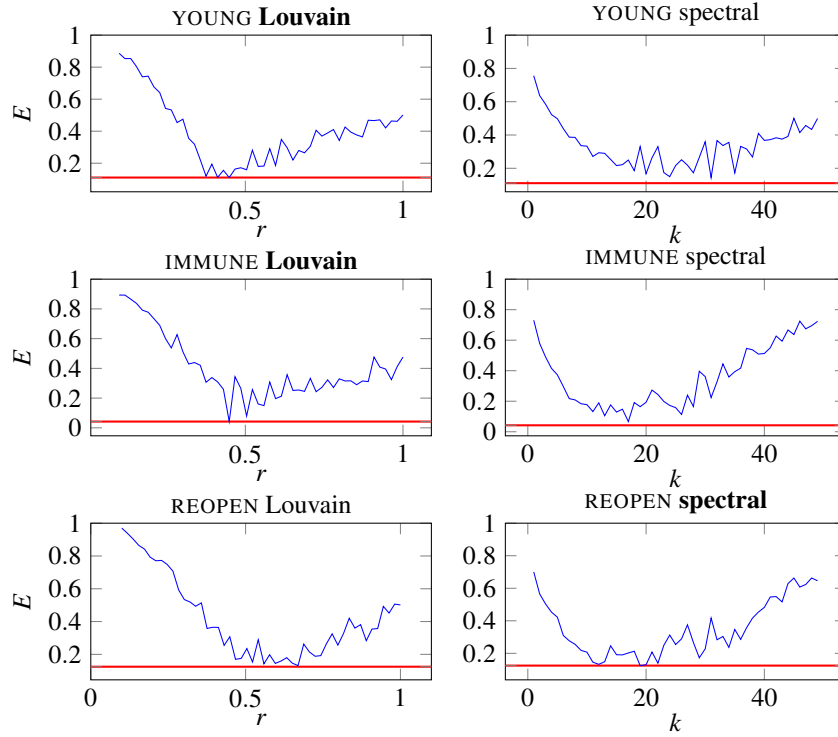


Figure 3. Parameter tuning for argument clustering with $E = 1$ for argument clusters of size 1. Repeats results from Figure 3 from the main paper, now showing the best score (red line) obtained by Louvain for YOUNG and IMMUNE, and spectral for REOPEN.

D.2.2. $E = 1$ vs $E = 0$ for single member clusters

We also experiment with setting $E = 0$ for argument clusters of size 1 (i.e., clusters containing only a single key argument). The results are displayed in Figure 4, overlaid over the previous results where $E = 1$ for single-member clusters (Figure 2). As expected, error is low when a large number of clusters are obtained by each method (low r , high k). The optimal parameter settings chosen in our approach corresponds to the tipping point where E switches between low E to high E .

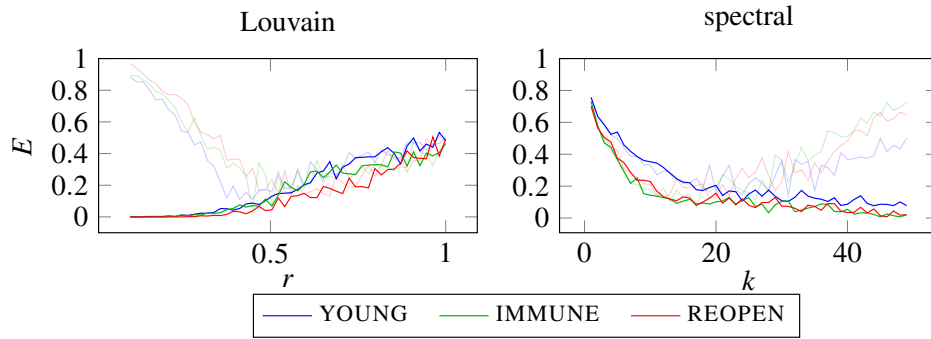


Figure 4. Parameter tuning for argument clustering with $E = 0$ for argument clusters of size 1. Results are overlaid on Figure 2.

D.3. Clustered Argument Stances

All arguments that were clustered in the second phase of HyEnA were extracted with a particular stance. The clustering method, either Louvain or spectral, clusters based on the obtained similarity labels. However, we can check the correspondence of all stances of the arguments within one cluster, as they should all match. Figure 5 reports the average stance errors per cluster for the three policy options. Stance error is defined as the proportion of stances that do not match the majority stance. In general, the error among stance labels is low; only in some cases mixed stances occur in the clustered arguments. Moreover, only in 5 out of 24 cases the non-majority stance occurs more than once, showing a high agreement between stances inside clusters.

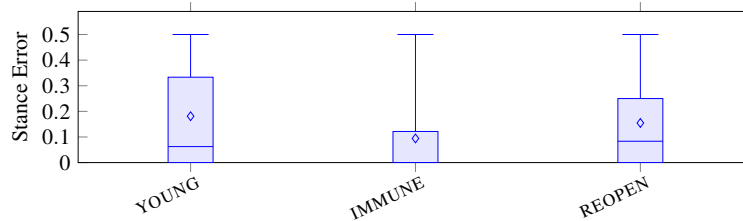


Figure 5. Stance error per final cluster of HyEnA. Overall, low error scores are achieved, indicating high stance correspondence inside clusters.

D.4. Annotator Reliability Analysis

Table 5 shows the inter-rater reliability (IRR) for four steps with overlapping human annotations. In the topic generation phase (Section 4.1), we use the intraclass correlation coefficient $ICC(3,k)$ [5] since it involves ordinal ratings. In the other three tasks, multiple binary labels are obtained for the same subjects. In these tasks, we use prevalence- and bias-adjusted κ (PABAK) [6], which adjusts Fleiss’ κ for prevalence and bias resulting from small or skewed distribution of ratings.

Table 5. Average (and standard deviation) IRR scores.

Task	ICC3k	PABAK
Topic generation	0.66 (0.14)	–
Topic assignment	–	0.81 (0.10)
Key argument consolidation	–	0.34 (0.03)
Key argument evaluation	–	0.40 (0.06)

The IRR for topic generation and assignment tasks are substantial. The IRR for key argument consolidation and argument evaluation are *fair* and *moderate*, respectively. We pose that the relatively low IRR scores of the latter two tasks are not shortcomings of the HyEnA method in itself. Instead, they demonstrate the complexity of language understanding, and the subtleties involved in interpreting and reasoning about arguments and opinions. Hence, hybrid approaches which use human insight are a key component for public feedback analysis. Uncovering these subtleties and making them explicit is a crucial task for enabling effective perspective taking [7]. This also justifies the need for a robust argument consolidation phase that integrates judgements from a range of interpretations.

D.5. Key Arguments

The key arguments extracted by HyEnA are shown in Tables 6, 7 and 8. The results for the ArgKP automated baseline are shown in Tables 9, 10 and 11. Tables 12, 13 and 14 show the results from the manual expert-driven baseline.

Table 6. All argument clusters from HyEnA for the option of *Young people may come together in small groups*.

Option	ID	Stance	Argument cluster
YOUNG	0	pro	⟨ Social contact is essential for development, It will be positive for support and acceptance, possitive for the psychological health of children, Young people have already suffered enough and got deprived of so many things like parties, holidays, sports. They are missing out on the best time of their lives, Young people's mental health will improve, Removes a lot of annoyance among the elderly, The lifting of this measure significantly reduces loneliness, while having minimal effects, Young people show more cooperation and thinking along when the way they live is taken into account, co they don't have to maintain distance ⟩
	1	pro	⟨ Going back to normality, Second wave, Following research results, this should be possible ⟩
	2	con	⟨ There's a limit to the restrictions, More measures lifted is good, As long as it can still be controlled ⟩
	3	pro	⟨ No risk of contamination , Young people have fewer contamination risks, It's not dangerous for the young people, The group is not at risk at dying of covid, Limited risk, large profit for that group, They're less likely to be contagious, and they're already together anyway. , Young people less infects ⟩
	4	con	⟨ Maintaining distance between your friends and family is easier than being locked down and deprived of the change to make a living ⟩
	5	con	⟨ Joggers don't maintain the distance and the effects of such behaviour are very small and negligible , Maintaining distance while exercising with each other is very difficult, It is dangerous for young people's health to don't keep the distance ⟩
	6	con	⟨ Risk of contamination, The infections will increase, The chances of the second peak of corona virus is too high, The risks are too large, The numbers of the infected have peaked following the holidays, Does not solve the risk of contamination, Unnecessary risk, Who has better immunity system will live, who not will die ⟩
	7	pro	⟨ Economy is more worth then the young ones, The economy will improve and companies won't go bankrupt, They still go to the pub, Life has to go on regardless of the situation, Young people would be happy about going out and meeting friends ⟩

Table 11 continued. All argument clusters from HyEnA for the option of *Young people may come together in small groups*.

Option	ID	Stance	Argument cluster
YOUNG	8	con	⟨ Exceptions should be considered, Because this cannot be maintained, and it is already violated everywhere, We should be cautious with making big changes to the regulations because it might cause us damage, Entertainment/Events give opportunities to break rules, with this option no longer risk of breaking rules ⟩
	9	con	⟨ People should reasonably decide the distance to maintain, They wouldn't switch between 1,5m distanz with old ones and young ones, they would always be nearer. , People will be more willing to meet and they will do it in larger groups which will enable the spread of the diseases, It is impossible to tell the exact age of people or gauge their immunity, Regional measures will cause problems because people commute between cities. ⟩
	10	pro	⟨ This measure will not be respected, The average Dutchman is too stupid to control themselves when out among people, It is impossible to stop it either way, They don't do it anyway regardless of the rules, People are not responsible enough for the measure to be dropped, They didn't keep the distance before, It is too difficult to follow this rule ⟩
	11	con	⟨ Important measure to archive immunity, Nursing homes can open up only if the measures are followed, Treating all people equally and not just the young ones ⟩
	12	con	⟨ Excessive measure, It saves a lot of tax for the police because they won't need to observe young people so closely, It is not proven yet whether this would be a good option ⟩
	13	con	⟨ To many young ones would gather ⟩
	14	con	⟨ One rule for all, The young people can contaminate others, Too early ⟩
	15	pro	⟨ Many people already don't do the 1,5m distance, Less victims if they use 1.5 meters at home with family members ⟩
	16	con	⟨ Lack of control, Easing encourages spread, Every life is worth more than the economy, Netherlands has more than enough resources to at least keep its head above water for a considerable time ⟩
	17	pro	⟨ Only the sick people should stay at home, the same as with the regular flu ⟩
	18	pro	⟨ Young people can study again and learn together, Children can go easier to school, The schools will be open soon anyway, Young people want to see and socialize with people again, Alternate the students that go to school and the other half attend classes at home ⟩
	19	con	⟨ People will spread the virus more quickly as they will feel more willing to meet in large groups ⟩

Table 7. Argument clusters from HyEnA for the option *All restrictions are lifted for persons who are immune.*

Option	ID	Stance	Argument cluster
IMMUNE	0	pro	⟨ it is fair to give immune people freedom of movement ⟩
	1	pro	⟨ could lead to a second peak in cases, These measures are easier to follow compared to other measures, This is a relatively easy measure to take, Public transport use would be easier ⟩
	2	con	⟨ People who still need to follow restrictions will be less likely to when others are not, Immune people would have advantages over the non-immune, and this is unfair, could be seen as discrimination, Everyone should be subject to the same set of rules/restrictions. , Complacency will make it harder for individuals to follow the rules, Young people seem to be getting an advantage over older people ⟩
	3	pro	⟨ Restrictions are unnecessary for people who are immune, Immune people should not be constrained ⟩
	4	con	⟨ Hard to maintain and/or implement, Too little research has been done, It is difficult to control, People can lie if they've contracted the virus ⟩
	5	pro	⟨ People will be able to meet with friends and family members again, It will allow things to get back to normal, People will be happier if they're allowed to go outside, People will be able to see family again, making them happier. , Family can visit each other more often, There will be solidarity between groups and regions, It is fair to give people back their freedom, People will be less lonely and depressed, People want to see their families again, and this measure allows it ⟩
	6	con	⟨ it is unclear if it will be helpful or will make things worse, ICU beds will become more crowded, It's still too early to relax ⟩
	7	con	⟨ It is hard to tell if people are truly immune, Not enough is known about the coronavirus yet, There are too few opportunities to test it, You can't tell who is immune and who isn't, One can lie about having or not having the virus ⟩
	8	pro	⟨ Current restrictions do not really provide any safety, This measure can have a negative effect on society ⟩
	9	con	⟨ It is not clear how people will be able to prove that they are immune, It is hard to know at a glance if someone is immune or not and this will allow some people to fake immunity, there could be immune people with other factors that make them vulnerable, immune people are no longer infective, People who are immune are not dangerous to others, Immunity has not been proven ⟩
	10	con	⟨ will funnel people in certain areas, Risks of transmitting the virus in gatherings ⟩
	11	con	⟨ Infection numbers are still increasing, It risks causing a spike in case numbers, Could lead to the misunderstanding that the situation is safe, Lifting restrictions will cause another wave of Covid, Lifting restrictions will cause people to stop following other rules related to Covid like social distancing. , Too much risk of another spike in cases, By taking this measure, health care would become very pressured ⟩
	12	con	⟨ Infections and morality will increase ⟩
	13	pro	⟨ Advantages to the economy from having immune people working again, This will be beneficial to the economy, People in high-risk of contact jobs will be allowed to return to work, Lifting restrictions will cause economic and social damage. , Lifting restrictions will allow people to feel like things are returning to the pre-Covid normal. , People can go back to work, People who work in contact professions can go back to work, Immune people are, well immune, and can help getting the economy back up ⟩

Table 8. All argument clusters from HyEnA for the option of *Re-open hospitality and entertainment industry*.

Option	ID	Stance	Argument cluster
REOPEN	0	pro	⟨ This will bring improvement in employment rate, This will improve the economy, This will help these industries recover, to support these sectors and to entertain and please us all, Killing the industry, This helps the economy ⟩
	1	con	⟨ will end up in another confinement, will end with a spike of infections, It is too early, There are less cases now than before ⟩
	2	con	⟨ The difference is we must first protect ourselves from this sickness to then adapt, This will help people satisfy their cravings, People will not benefit a lot from this, This can help people create social interaction and build resistance against COVID ⟩
	3	con	⟨ Leads to more COVID cases , Leads to better moral While keeping Covid cases down, If people die business will still suffer , Things aren't normal yet, Keep sick people away, This will bring more new cases and deaths ⟩
	4	pro	⟨ This can be done only on open spaces, It's already being done in other countries, There are more important industries that needs to be re-opened. , This will help people earn enough to support basic necessities, Tests can be previously made ⟩
	5	con	⟨ will gather a lot of people together, Better moral less infection , This will bring about chaos and lack of control ⟩
	6	con	⟨ These industries are very risky, Risk of spread increases significantly, Catering is a distance of 1.5 meters impossible which leads to great chance of contamination, This increases the chances for the virus to be spread ⟩
	7	pro	⟨ will decrease the number of people with breakdowns, will decrease the contact between people, Keeping group small helps ⟩
	8	pro	⟨ will increase the attendees in the shows, will be controlled environment, With the necessary restrictive measures, cultural events must be able to be visited again as they are an important part of human life, Workers are well protected ⟩
	9	pro	⟨ No evidence that the lockdown works, A distinction should be made, some contact professions are basic service and others are not, Restriction of liberty is a violation of human rights ⟩
	10	pro	⟨ Excited to do things as before for preserving mental health, This will ensure freedom for the people, In order to save people 's lives, we should be very careful and not relax too quickly, To support the churches and meet fellow believers again and pray and sing together ⟩
	11	con	⟨ It's not worth getting people sick, It's not safe yet , These are not vital industries ⟩
	12	pro	⟨ People need to let out pressure , People are tired and bored , Culture and entertainment is important in life, This will make people feel better ⟩
	13	pro	⟨ It will help everyone tremendously, This will help people go back to work, This will motivate people to be more active and healthy ⟩
	14	pro	⟨ Need freedom, It is best to know more of the virus before reopening these industries, This can be done following certain conditions, This will support small businesses recover ⟩
	15	pro	⟨ This will empower the people to be more responsible ⟩
	16	pro	⟨ Cannot be maintained, These places can't be maintained ⟩
	17	pro	⟨ It is easy to maintain social distancing in these industries. ⟩

Table 9. All arguments from ArgKP for the option of *Young people may come together in small groups*.

Option	Stance	Arguments
YOUNG	pro	in the long term, this measure is not sustainable in any case
	pro	Low risk group. Easing also gives more space for parents/families.
	pro	if it is not necessary then it is desirable. Also saves on enforcement
	pro	Easing at 1.5m may provide better motivation to comply with other measures
	pro	Youth has the future, it pays a lot for what it 'costs'
	pro	This is hard to maintain. Let's put time into more urgent matters.
	pro	young people are not going to last , a lot of fighting in home situation
	pro	Young people need to support the economy again by getting to work
	pro	Young people need freedom, encourage their own responsibility
	pro	Schools can open 100% again, so parents can also work 100% again
	pro	Can't be stopped. Maintaining this leaves society in a state of cramp.
	pro	Up to the age of 18, this must be the responsibility of parents.
	pro	Relatively little extra pressure on care. Easing this measure benefits education.
	pro	they already had a lot of trouble with it, making it better official
	pro	Untenable for that group, but appeal to solidarity with at-risk groups
	pro	young people do not have the full support to risk
	pro	Help for parents to work better at home
	con	Immunity has not yet been proven. Young people can also transmit the virus.
	con	The rules must remain uniform, otherwise there will be confusion
	con	Young people are better at fighting the Coronavirus
	con	see previous answer Health is for economic importance
	con	young people don't care much about the same problem
	con	We must all stand in solidarity. Moreover, enforcement is easier
	con	Groups with relatively small economic impact if the measures continue to exist for longer.
	con	That way you distinguish between people. This is not advisable for maintaining support.
	con	Young people can easily transfer. No physical/mental distinction between people.
	con	no exceptions for subgroups. Together we get corona under control.
	con	In fact, my motivation is: Equal monks, equal caps.
	con	I don't want to be responsible for the deaths of fellow human beings.
	con	Risk hedging in the near future. Adds nothing
	con	because I am not convinced that well-considered visionary decisions are now being taken
	con	Companies are always at the forefront. Now health comes first No generational differences
	con	Everything is making choices
	con	based on the effects in the explanatory statement, I make that choice.

Table 10. All arguments from ArgKP for the option of *All restrictions are lifted for persons who are immune.*

Option	Stance	Arguments
IMMUNE	pro	Partly rekindling the economy Better availability of healthcare staff Less protective equipment needed
	pro	that can be used in crucial places
	pro	If you maintain it, I think this is a logical choice.
	pro	Positive effect on loss of income for large group of people.
	pro	Why restrict people's freedom when there's no very urgent reason for it?
	pro	No, it just has to be suffering.
	pro	people are perfectly capable of using their common sense
	pro	The psychological benefits seem much greater than the physical disadvantages.
	pro	they can be deserving of people who are sick
	pro	You can decide what you want. Some feel deprived of their freedom.
	pro	This makes travelling in public transport easier, for example
	pro	These people can therefore reduce the uneaten of the elderly
	pro	Everyone has to be free, but living in a dictatorship very sad
	pro	Survival of the fittest. Reward is in order
	pro	That should be possible n arithmetic could not predict a future
	pro	This seems like a good start to moving for the new world name corona virus
	con	Immunity has not yet been proven. Young people can also transmit the virus.
	con	Immunity has not been established Opening certain provinces gives much more travel
	con	Creates inequality that is not good for social cohesion. Possible source of polarization.
	con	this reduces the willingness of the rest of the netherlands
	con	Too much risk people don't have a size if they are allowed again
	con	Because young people don't stick to it now so it won't matter much
	con	see previous answer Health is for economic importance
	con	In my opinion, the selected items are less urgent than the other
	con	This gives a high degree of inequality within the population
	con	It's way too early for that. R values must remain well below 1
	con	Don't reward groups for already having a problem with the rules.
	con	Because we want to live a normal life again
	con	no exceptions for subgroups. Together we get corona under control.
	con	Enforceability is complicated, keeps simple rules. Moreover, these measures undermine solidarity.
	con	This is uncheckable, you have to show proof everywhere.
	con	because I am not convinced that well-considered visionary decisions are now being taken

Table 11. All arguments from ArgKP for the option of *Re-open hospitality and entertainment industry*.

Option	Stance	Arguments
REOPEN	pro	Catering under certain conditions. entertainment as late as possible
	pro	Empower citizens' own responsibilities
	pro	I think those at high risk can be advised to avoid hospitality.
	pro	Hospitality but not entertainment. Catering reasonably similar to shops.
	pro	Only when you're sick do you stay at home, otherwise you don't
	pro	visitors are usually under 50 years of age, can handle this
	pro	Especially lower risk groups use these facilities.
	pro	Everyone can decide for themselves whether they want to go here.
	pro	people are perfectly capable of using their common sense
	pro	People know how to do this. Sufficiently alert to allow this.
	pro	restriction of liberty is violation of human rights
	pro	Make sure the drug is widely available, then the percentages will be even lower
	pro	Who else is going to pay the extra care costs?
	pro	Have seen so many good ideas on media to open responsibly
	pro	Income is also important. Over-50s don't have to participate.
	pro	These companies are also on the rise.
	con	lifting measures northern provinces suffer from hospitality migration within the Netherlands
	con	These options can cause other problems, are uncheckable or easy to by-pass.
	con	Too much risk. People will then travel to those regions.
	con	Risk of spreading is far too great. Measure 1.5 meters is impracticable
	con	No distinction between areas in NL Entertainment is less important.
	con	Too dangerous for too little added value.
	con	Somewhere we have to start slowly with normal life again, but with limitations.
	con	Equal treatment of the population
	con	I believe that public support for safety will be greatly reduced.
	con	People are well able to weigh up themselves
	con	people have common sense
	con	A personal choice is not one of the government's.
	con	This is uncheckable, you have to show proof everywhere.
	con	because I am not convinced that well-considered visionary decisions are now being taken
	con	Restaurants also cause addiction damage

Table 12. All arguments from the expert-driven manual analysis for the option of *Young people may come together in small groups*. Arguments are **mapped to** argument clusters from HyEnA, showing the cluster ID taken from Table 6.

Option	ID	Stance	Arguments	Mapped to
YOUNG	0	pro	Young people play a minor role in the spread of the virus and their risk of getting sick is low	3
	1	pro	Social contact is relatively important for young people (to develop themselves)	0
	2	pro	For young people it is difficult not to violate the rules	10
	3	pro	Reduction of problematic psychological symptoms	0
	4	pro	Reduces the pressure on parents	–
	5	pro	Possibility to build up herd immunity	11
	6	pro	Increases support among young people for other lockdown measures	1
	7	con	Constitutes age discrimination which results in a dichotomy in society	14
	8	con	Measures are difficult to enforce. Young people will also get in contact with other people	8

Table 13. All arguments from the expert-driven manual analysis for the option of *All restrictions are lifted for persons who are immune*. Arguments are **mapped to** argument clusters from HyEnA, showing the cluster ID taken from Table 7.

Option	ID	Stance	Arguments	Mapped to
IMMUNE	0	pro	These people pose no danger to their environment	3
	1	pro	These people can keep society and the economy going again	13
	2	pro	It is pointless to demand solidarity from these people if they are already immune. Doing so will lead to fierce protests	8
	3	con	Tests for immunity are not foolproof, and this increases the risk of new infections	11
	4	con	Creates a dichotomy in society. People who are not immune can get annoyed by the behaviour of those who are allowed to resume normal life	2
	5	con	Difficult to enforce	4
	6	con	Potential confusion as immunity is not outwardly apparent	7

Table 14. All arguments from the expert-driven manual analysis for the option of *Re-open hospitality and entertainment industry*. Arguments are **mapped to** argument clusters from HyEnA, showing the cluster ID taken from Table 8.

Option	ID	Stance	Arguments	Mapped to
REOPEN	0	pro	This is good for our economy and business	0
	1	pro	It is good for people's well-being	12
	2	pro	This relaxation option will increase support for the continuation of the other measures	–
	3	pro	It is enforceable	4
	4	pro	People can take responsibility for themselves by staying away if they wish	15
	5	pro	We should preserve our cultural heritage and cannot risk bankruptcies in the cultural sector	12
	6	pro	Keeping these businesses closed is too big of a sacrifice for young people	–
	7	pro	In this way, we can build up herd immunity	–
	8	pro	If the hospitality industry is not re-opened people will do other things to relax which is also risky	9
	9	con	Risk of too many people gathering together, which helps to spread the virus	3
	10	con	It is not necessary at the moment	11
	11	con	When alcohol is consumed, people are more likely to underestimate risks and are less likely to comply with distancing measures	–
	12	con	Opening up the hospitality and entertainment sectors should only be considered in the next phase if it appears that other adjustments have worked	14
	13	con	Hospitality industry has a bad impact on society. Please keep it closed	16

References

- [1] van der Meer M, Liscio E, Jonker CM, Plaat A, Vossen P, Murukannaiah PK. HyEnA: A Hybrid Method for Extracting Arguments from Opinions: Supplementary Material; 2022. <https://liacs.leidenuniv.nl/~meermtvander/publications/hyena/>.
- [2] Chai C, Li G, Li J, Deng D, Feng J. Cost-effective crowdsourced entity resolution: A partial-order approach. In: Proceedings of the 2016 International Conference on Management of Data; 2016. p. 969-84. Available from: https://dl.acm.org/doi/pdf/10.1145/2882903.2915252?casa_token=Nb3MPRW1PxcAAAAA:uKZ1FJO8sPHloM3UC2_cJR7EaL2Bhx7dEP1dXe6pyjUGitjC7AMH1doe0Radf6YqBN4GCUwU1M3UCg.
- [3] Bar-Haim R, Kantor Y, Eden L, Friedman R, Lahav D, Slonim N. Quantitative Argument Summarization and beyond: Cross-Domain Key Point Analysis. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2020. p. 39-49. Available from: <https://aclanthology.org/2020.emnlp-main.3.pdf>.
- [4] Gretz S, Friedman R, Cohen-Karlik E, Toledo A, Lahav D, Aharonov R, et al. A large-scale dataset for argument quality ranking: Construction and analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34; 2020. p. 7805-13.
- [5] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*. 1979;86(2):420.
- [6] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy*. 2005;85(3):257-68.
- [7] Chen S, Khashabi D, Yin W, Callison-Burch C, Roth D. Seeing Things from a Different Angle: Discovering Diverse Perspectives about Claims. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); 2019. p. 542-57. Available from: <https://www.aclweb.org/anthology/N19-1053.pdf>.