brac
institute of
governance &
development

BIGD WORKING PAPER

NO. 67 ■ NOVEMBER 2022

Sentiment of Bangladeshi Residents Toward Covid-19 Lockdowns: Qualitative Analyses of Open-Ended Responses in a Large Panel Survey

Daryl Collins, Derry Moore, Pravarakya Reddy Battula,
and Avinno Faruk

GOVERNANCE AND POLITICS SERIES

Sentiment of Bangladeshi Residents Toward Covid-19 Lockdowns

Qualitative Analyses of Open-Ended Responses in a Large Panel Survey

Daryl Collins, Derry Moore,
Pravarakya Reddy Battula, and Avinno Faruk

BIGD Working Paper
No. 67 | Governance and Politics Series

November 2022

BRAC Institute of Governance and Developement (BIGD)
BRAC University

Abstract

This paper explores public sentiment of Bangladeshi residents concerning the lockdowns imposed by the Bangladeshi government in 2021 in response to COVID-19. Through open-ended question design and analyses of natural language using NLP and sociolinguistic techniques, we show detailed, nuanced sentiment as well as common themes and discussion these sentiments are seated in. Additionally, using a range of discursive analytical measures, we explore the interactions between enumerators and participants in live survey conditions, providing alternative methods to and potential field guidance for enumerator survey methods.

Contents

1. Introduction

Across the world, government-imposed lockdowns comprised a consistent strategy for limiting surges in COVID-19 cases. The impacts of these lockdowns were manifold and had significant impacts on the livelihoods of residents. Understanding the sentiment and how people feel about these lockdowns can better frame an understanding of impact on livelihoods.

For this project, Decodis partnered with BRAC Institute of Governance and Development (BIGD) to explore questions related to public sentiment of government-ordered lockdowns in Bangladesh. We explore these questions through an open-ended question design implemented within a large panel survey conducted by BIGD with 3000 participants. By leveraging Natural Language Processing (NLP) and sociolinguistic techniques, we are able to better ascertain granular-level, "on the ground" sentiments of these Bangladeshi residents.

The panel study took place through live interviews over the phone. Enumerators recorded themselves asking the questions and the respondents' answers. Because of technical difficulties getting the enumerators to record the responses (only 750 were able to record) as well as many having inaudible responses, the final sample for the three questions was just under 400 Future studies may explore ways Decodis's IVR methods can address these issues. Additionally, we conduct further analysis in this project on the discourses of enumerators and the consequent potential effects on data.

2. Method

2.1. Question Design

In designing questions, Decodis is careful in wording and instruction with intention to increase how long respondents talk for, the depth of their answers, and their engagement in the survey as a whole. We use the following principles of design in wording and constructing questions.

2.1.1. Questions are Interactional

Questions are interactional in nature whereby the question – answer pair is, under normal discourse proceedings, distributed between two conversants. Questions imply answers. In fact, "the force of a question is... to elicit a particular kind of answer" [emphasis added]. (Levinson 1979) When asking a question, we are, necessarily, begging for an answer. Beyond the linguistic interaction, we are also invoking a social interaction. One that is subject to all the social pressures and forces of any other. As such, context becomes immeasurably important to consider (for further reading see Levinson 1979, Athanasiadou 1991, or De Ruiter 2012).

We must take into consideration the contexts of those asking the questions, those responding, as well as the interactional context itself. Askers will ask questions with preconceived notions of possible answers and respondents will answer with sets of preconceived notions of the askers intentions in asking. We must then attempt to consider what those may be ahead of time. In so doing, we may consider power dynamics between parties involved and the ways these may hinder, hamper, or, instead, help get the kinds of responses we seek.

2.1.2. Respect for Participants

Connected to considering context, we must always ask questions in ways that respect persons. We should not talk down, questions should be asked in an understandable way, and participants should be given the freedom and space to answer in whatever way they choose.

2.2. Question Choice

A critical concern of question design in surveys and questionnaires is the give and take between efficiency of information gathering on one hand and the potential breadth in answers on the other. Directed, close-ended questions favor the former ("What do you do for a living?") while open-ended questions favor the latter ("Can you tell me about your job?"). Close-ended questions allow for quick coding and clear coding schema. Open-ended questions, however, allow for richer, more varied responses and allow for the possibility of more discovery.

Because a core component of Decodis's philosophy is to bring individual voices to the forefront, to give power to the respondents' voices, we rely on open-ended question design, seeking ways to allow respondents to answer however they feel for as long as they want.

The questions chosen for this project reflected Decodis's best guess at where we could most add value to the panel surveys. Therefore, we focused on questions that, we felt, people would have strong opinions about and would therefore want to talk more about. The three questions are presented below in Table 1.

Table 2.1. Questions selected for study

	Question (English)	Question (Bangla)
Q72.3	What do you think of the government's decision to impose a lockdown to contain Corona?	Coronar somoy sorkarer lockdown deyar siddhanto ke apni ki mone koren? (sob cheye gurottpurno ekti uttor niben)
Q79.9	Aside from food items, what type of relief would enable your household to return to a better condition?	Khaddo-drobbo chara ki ki dhoroner sahajjo pele apnar poribarer arthik obostha ager obosthai fire jabe bole mone koren? (gurottopurno sorboccho 3 ti)
Q79.10	Tell me any hesitancy you feel about asking for assistance/relief?	Tran/onnanno sohojogita chawar bishoye apnar ki ki dhoroner shongkoch kaj kore ta somporke amake bolun. (ekadhik)

2.3. Data Collection

Data was collected during the survey phone call by enumerators hired by BIGD. This team of enumerators recorded sessions of the survey, recording both themselves asking questions and respondents answering them. A total of more than 5000 responses were recorded.

When asking the open-ended questions for this study, enumerators were instructed not to read out the previously formulated options from the instrument, but to select the nearest answer given the respondent's response. Additionally, they were instructed to not provide follow up information, but to simply ask respondents to reply to the best of their ability and understanding. This would ideally minimize the influence of the enumerator in the respondents answer to the open-ended questions.

A critical part of our methodology includes the analysis of audio, more specifically the acoustic signal. "Acoustic signal" in this context refers to the speech produced by participants as measured through the physical properties of the sound waves. As speakers vibrate their vocal cords and shape their mouths, they actively manipulate the sound waves they project, this "acoustic signal" is then interpreted by listeners as speech. This acoustic signal can be examined in all sorts of ways for a multitude of purposes. In this project, we primarily use the acoustic signal as a window into speakers' emotional states. Unfortunately, the reliability of results in acoustic measurements drops significantly with poor audio recording quality. Through testing, we determined that audio with insufficient loudness dramatically reduced the accuracy of pitch measures, a critical measure to our analyses. Therefore, to prevent wasted effort or misleading measures, we established a minimum intensity value against which every audio was assessed. If the average intensity across an audio did not meet that minimum intensity threshold, then it was determined to be of too poor a quality to be usable. About 750 out of the 5,000 responses were determined to be potentially useable.

It is important to note that this study was a pilot and, therefore, there was limited technical preparation for recording responses.

2.4. Transcription and Translation

For these 750 audios, we enlisted a team of transcribers/translators (henceforth, simply transcribers). Transcribers produced the written text of both the enumerators and the participants as well as aided us in our interpretation of expressions or speech that required native speaker intuition and input. In addition to transcribing and translating the responses, these transcribers also annotated the time stamps for each "utterance" of a response.

"Utterance", here, can be thought of as a spurt of speech. They are complete "speech-moments", usually expressing a single thought, and usually bounded by pauses. Utterances usually correspond to sentences, but, depending on the nature of the language, the participants, and the conversational context, may be less than a sentence or multiple sentences. Consider, for example:

A: "What do you think about the lockdowns?"

B: "Bad"

Here, Speaker B does not use a full sentence. However, they have completely answered the question and expressed the thought in full. We would not necessarily expect – even though we may want – further expression from Speaker B.

Similarly, a context in which a speaker shifts thoughts midstream would likely constitute an utterance since the thought is "completed" discursively (though not in actuality):

B: Well, I was on my way to…wait, why do you ask? (where the ellipsis "…" would constitute the utterance boundary).

As transcribers worked through these audios, they found many more to be inaudible, garbled, or otherwise of insufficient quality. Resulting in a final sample between 300 and 400 responses, depending on the question. Because questions were recorded in different sessions and enumerators were inconsistent in recording practice, the final number of "good" audios varied by question.

2.5. Data Analysis

To analyze the data collected in this study, our team used tools and techniques from Natural Language Processing (NLP), Sociolinguistics and cluster methods.

2.5.1. Natural Language Processing (NLP)

2.5.1.1. Topic Modeling

Topic modeling is a method whereby a set of texts (here, the responses to questions) are categorized based on the sets of common terms used throughout those texts (for reference see Vayansky & Kumar 2020). This method does not inherently show any connection between the sets of words (i.e. it cannot "define" the topic) beyond their simple, regular co-occurrence. Themes and topics are more than just a set of content words, but rather come about through natural language use and full expressions. As such, it is up to the researcher to examine the word-sets to better understand their co-occurrences and to, subsequently, define a set of topics/themes. We accomplish this by first creating a subset of data containing a small sample of responses for each word-set. We then examine these responses for common themes as laid out by the word-set. Through this exercise, we (a) may decide that multiple word-set groups should be combined and (b) label the word-sets with a topic/theme name.

2.5.1.2. Topic Classification

The next step in our process is to attempt to improve the topic labeling by creating a more specific labeling system. For this, we use a method of Topic Classification whereby specific words or expressions are coded for each topic. Each response then is queried against these codes and annotated accordingly. This method allows for a much more precise method of annotation and affords us the insight into varied usage of certain words and topics mentioned in a response. While in Topic Modeling each response is uniquely assigned to a particular set of commonly co-occurring words, here in Topic Classification, each response may be tagged under multiple topics.

2.6. Sociolinguistics

It is no surprise that the words people use do not tell the whole story. There is a wealth of meaning built into how people say the things they say as well. How a participant thinks and feels about issues and topics critically impacts our understanding of their perspective and is, therefore, of significant importance and interest to this project.

A large body of research has explored the various linguistic reflexes of emotion across subfields of sociolinguistics, psycholinguistics, and computational linguistics. Drawing on topics of research such as Attitudes, Stance, and Speech Emotion Recognition, we seek to tie features from the acoustic signal to the emotional/attitudinal frame of individual speakers. By adding this layer of analysis we are better able to place our analyses from the perspectives of the participants, allowing for clearer shading of the discussion of topics and the ability to discover insights obscured by text alone.

Across the literature a large set of acoustic features are tied to emotion analyses (see Scherer 2003). While there is some debate on the particular outworkings of these features across cultures, there is some evidence for a certain amount of universality to certain emotion fundamentals (e.g. activation) and their acoustic correlates (e.g. intensity) (Feraru et al. 2015). Pitch (voice height) and Intensity (loudness) almost always come up as relevant features across the literature. Therefore, we rely heavily on those measures. We also use derived measures of pitch variation (PDQ) (Hincks 2004) across responses as well as the length of responses (duration).

To extract these measurements, we use an acoustic analysis program called Praat. Using a script, we step through each utterance within an audio recording in 50msec windows, extracting the average pitch and intensity within each window where pitch is detected, ignoring any pauses or non-speech. We then normalize those measures using Lobanov Z-score normalization (Lobanov 1971), normalizing to the average (mean) pitch value for each individual across the entirety of their responses (all three questions). In this way, we can reduce the effects of many confounds to speech signal variation like gender, given that women generally have higher pitch than men, or idiosyncratic difference, since some speakers may tend to speak softer than others. These normalized values are then averaged across the utterances to obtain response-level measures.

2.6.1. Emotional models of interpretation

In reading the speech signals for emotional information, we draw on models of interpretation from various sources (Scherer 2003, Jiang and Pell 2001, Ayadi et al. 2011, Fish et al. 2017, inter alia) as well as interpretations provided to us by translators and native speakers. First, it is important to note that the speech signal interpretations (the emotional analyses) are layered on top of the textual analyses. That is, they are, to some degree, beholden to the text. Perhaps more precisely, they are interpreted within the context of the text.

So, taking the text into consideration, the speech signals are generally interpreted according to two emotional fundamentals (Laukka et al. 2005): activation, and valence. Activation refers to the “strength” of the emotion, how emotive the respondent is being. It is the difference, for example, between “happy” and “elated”, “frustrated” and “angry”. Valence refers to the positivity or negativity of an emotion. This being the difference between, “happy” and “angry”, for instance.

Generally, intensity and pitch variation most closely correlate with activation. Therefore, with increased intensity and/or increase pitch variation comes increased activation or emotiveness. We usually read these kinds of measures as a heightened “engagement”, meaning that participants are interested and engaged in answering the question.

Valence usually coincides with pitch values. Typically, the higher the pitch, the higher (more positive) the valence. This doesn’t always hold true, however, and generally we place more weight on activation interpretations of speech signals over valence. However, taking the word choice together with pitch can aid in valence interpretations.

In addition to these measures, we also take duration, the length of time speaking, to be meaningful in responses. Generally, speakers engaged and interested in their response will speak for longer. Alternatively, speakers giving very short, simple responses are often expressing a sort-of “canned” response, whereby they are simply answering in the way that they believe they are supposed to. These speakers are often less engaged (lower activation) in their responses and give very “empty” responses. Therefore, we will often remove these kinds of responses from further analysis.

2.6.2. Discourse Analyses

In addition to socio-acoustic analyses, we also draw on other sociolinguistic-behavioral analytical models including discourse analysis. We should make clear here that in linguistic tradition “discourse analysis” can have different meanings usually designated through capitalization: “discourse” vs “Discourse”. In this project we use both of these analytical frameworks in different exercises.

In one sense, discourse analysis involves analyses of linguistic-pragmatic behaviors substantively seated in discursive contexts (for reference see Levinson 1979). Necessarily, these sorts of analyses examine language in interaction between two or more interlocuters and the ways in which these speakers navigate social-pragmatic forces within that discourse context. These include behaviors such as turn-taking, terms of address, greetings and farewells, or other such phenomena. These analyses take into consideration the pressures and constraints of the discursive context on the behaviors themselves and the ways in which speakers navigate, push back, or reinforce these factors.

In this project we use these sorts of analytical frames to examine the interactions between enumerators and participants, examining the ways certain discursive behaviors affect or interact with social power dynamic within this interaction and how those behaviors may affect the data obtained.

Second, Discourse analysis (note the capitalization) instead examines what speakers say from a broader socio-cultural perspective, appealing to notions of “Discourses” that people at large engage with (see Johnstone 2017). Within this frame, we address linguistic behaviors within a broader social context. In this perspective, linguistic expressions do not exist in isolation. Rather, everything a speaker says connects to other things that speaker has said or heard. In this way, linguistic expression is inextricably woven with Discourse, ideological and socio-cultural notions that pervade a society. We examine, then, the ways speakers reinforce – or rebuff – these

Discourses through the ways they interact with them: the words they choose, the frames for their arguments, the implications of their expressions.

Importantly, we are not following a tradition of Critical Discourse Analysis (CDA) (see Blommaert and Bulcaen 2000). CDA turns a critical eye upon Discourses, examining the ways in which these Discourses reinforce power differentials between the privileged and the disadvantaged. This sort of frame and perspective would be disingenuous in the context of this project and is therefore not used here.

For this project, we use Discourse analysis to analyze shared topics of discussion across participants. We both identify shared Discourses as well as address how speakers engage with them.

To read more on these sorts of methods and perspectives, see Edlesky 1981.

2.7. NLP + Sociolinguistics: Clusters

As indicated above, the sociolinguistic analyses take the textual analyses (NLP results) into consideration in determining the speech signal interpretation. To do this, each response is first annotated according to the topics it contains, as well as specific speech signal measures. At the response level, we record the average normalized pitch, intensity and PDQ across the response as well as the duration.

Using these data, we use K-means clustering to group respondents together based on the topics they express and how they express them. We then pivot these data against each other, examining the clusters against their speech signals to determine a category of response. For example, we may find that, in general, respondents referring to topic A tend to speak with enthusiasm and engagement, while respondents to topic B express frustration or resentment. Additionally, we may find that speakers expressing topics fall into separate groups such that, for example, respondents talking about Topic C express either exuberance or dejection. As such, we reconfigure participant categories according to these clusters and speech signal interpretations.

3. Data, Results and Discussion

In the following sections, we will address key insights uncovered through the course of our analyses. These data may cross-cut questions. We will detail the nature of the relevant data along with our resulting analyses, comparing these results to those obtained through an enumeration process. Secondarily, because we see language as interactive, we also provide an analysis wherein we examine discursive behaviors of the enumerators, illustrating potential influences on data collection.

3.1. Attitudes to COVID-19 Lockdown

This section is based on 353 responses to the panel question 72.3: "What do you think of the government's decision to impose a lockdown to contain Corona?" Here, our first layer of analyses is a sentiment coding schema whereby responses fall into one of four categories: Positive, Negative, Both, or Neither based on what expressions respondents used.

Responses were coded into these categories using the NLP method Topic Classification, adapted to sentiment detection. We first examined subsets of the data by hand, then custom built a sentiment model based on those data to code the rest. This process is also iterative by design. So, once the model is sufficiently expansive, and the full data is coded, we select subsets of unannotated data to check if new expressions ought to be included in the model. We also select subsets of annotated data to check for mis-categorizations. The model is subsequently updated, and further iterations are performed if necessary.

Positive responses were characterized by expressions belying a positive attitude towards the lockdowns and were tagged based on words or phrases including (but not limited to) “good”, “agree”, “blessing”, or “not bad”: “No, what the Government is doing, they are doing it for good, and people.” Negative responses, on the other hand, express negative opinions towards the lockdowns through expressions such as “bad”, “problem” or “harmful”: “It’s harmful for poor people like us. Harmful.”; “No not a good decision at all. No, it’s bad for us.” If responses were tagged with both negative and positive expressions, they were categorized into a Both category. Inherently, these responses are longer and more nuanced – more on this later. For example, responses such as “It was a good decision for our health but bad in the aspect of earning money” or “It may be good if it is imposed, but we suffer in all kinds of way.” If a response was not identified as either positive or negative, it was classed as Neither. These responses may still express ideologies or opinions, but the sentiment of them is less clearly decipherable: “I don’t know about that. I live under Bangladesh Government. I follow the Government’s decision. I don’t know about the others.”

In this data roughly one third of respondents had positive responses, while negative responses accounted for about a quarter of responses. Another quarter of responses had both positive and negative expressions while just over a tenth expressed neither. Figure 3.1 below illustrates this division.

This result, already, is quite different from the enumerator coding. Enumerators had three options for coding: “Good” (44% of responses), “Good but bad for Livelihood” (12%), or “Others”(44%). The enumerator results show much higher positive responses than our analyses. Additionally, if we consider “good but bad for livelihood” as an equivalent to our “Both” category, the enumerator results show half as many in Both. In this coding scheme, there’s also no overt way for enumerators to count negative or, in keeping with a good-bad schema, “bad” answers. All non-good answers are subsumed beneath “Others”. This sort of schema obscures nuance in the results, as we show.

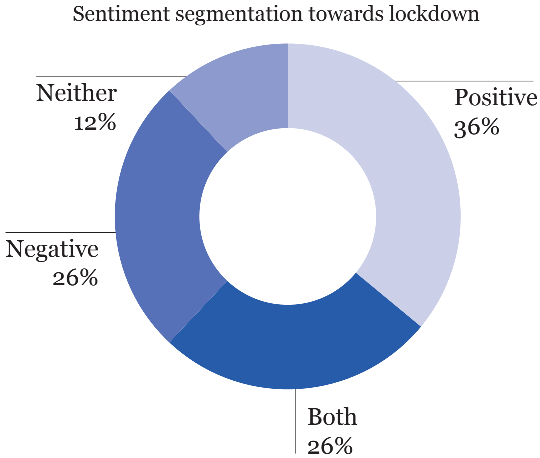

Figure 3.1. Segmentation based on sentiment of responses to “What do you think of the government’s decision to impose a lockdown to contain Corona?”

However, even if these issues were able to be solved with a different coding schema provided to enumerators, they don’t provide a great deal of insight beyond broad sentiment. So, we conducted further exercises to dig deeper into responses to add greater insight into how these respondents felt about the lockdowns.

3.1.1. Prevailing Themes on the Lockdowns

We first used Topic Modeling to search for common or prevailing themes within responses. Following this methodology, we found three persistent themes concerning the lockdowns that emerged in the data: “safety”, “the poor”, and “economic impact”. These themes accounted for 85% of all responses, meaning that most responses reference at least one of these themes.

3.1.1.1. Safety

These references, discuss the lockdown from a perspective of protection or preventing the spread of COVID-19, often in directly claiming “safety” as a primary reason for the lockdowns: “The government will do whatever they think and they always think about our safety first.” While both “the poor” and “economic impact” have some overlaps thematically, they are not the same. As a topic, “the poor” includes discussions of the effects of the lockdown on poor people specifically. On the other hand, “Economic impact”, may not specifically reference the impoverished, but rather is a topic of discussion centering on the shifts in economies. It is the difference between: “Poor people cannot eat when it is lockdown as they cannot go out to work and earn money.” and “No work means no income, means no food.”

3.1.1.2. Poor – First-Person Plurals as ingroup identifiers

The use of personal pronouns (or their grammatical analogs) performs a sort of “social deixis”, a positioning of the referrer in social distance to the referent (Hart 2010, Cummings 2005). By looking at the way respondents structure their references to “the poor” – by seeing whether they use first-person expressions or not (“I”, “me”, “we”, “us”) – we can infer whether the respondent self-identifies with a “poor” in-group (for reference see Duszak 2002, Hart 2010). That is, if a respondent answers, for example, with “We poor are starving”, the use of the “we” expresses the speaker’s personal experience to poverty (they refer to themselves) and further connects them to a collective experience (they use “we” not “I”). Other speakers, on the other hand, may use more “distanced” language (“Poor people can’t get food” or “They are starving”). We found that over two thirds (69%) of respondents who talk about “poor” self-identify as poor in this way (see Figure 3.2).

Respondents identifying as poor relative to total respondents talking about “poor”

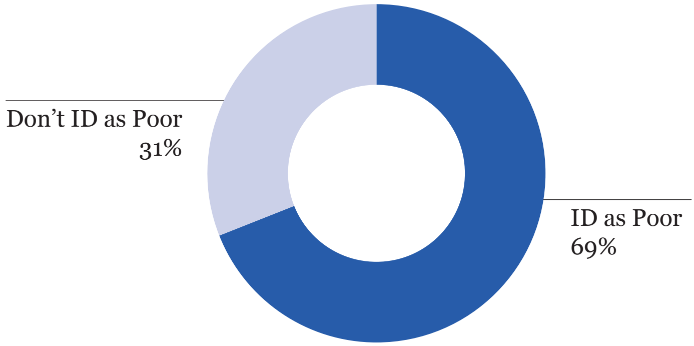

Figure 3.2. Proportion of respondents identifying as poor vs those who do not

Choosing to identify as poor or not carries potential social meaning beyond the surface. While the surface reading sees this as a metric for who is poor or not, or that those who don’t identify as poor are, themselves, not poor or not experiencing those issues, this may not be true. Some respondents may not self-identify as poor for a number of reasons. For one, there is a distancing effect. By distancing themselves personally from the status of “Poor” they are open to discuss the trials faced but without the potential embarrassment that may come along with that status. Additionally, this way of framing taps into a social discourse, indicating a generalized, more widespread concern surrounding notions of poverty. Another reason may be that these respondents are newly poor, becoming poor during the pandemic or because of the lockdowns. As such, they may not view themselves as one of “the Poor”. Respondents identifying as poor, however, may instead be recognizing their situation and reacting by appealing to a larger concern while verifying that concern through personal experience. Indeed, many of the responses we see in this sample seem to accomplish this: “poor people, like us, are suffering”.

Following this thread, we find that, over two-thirds (69%) of respondents discussing "the poor" use this sort of self-reference and are consequently tagged as identifying as poor. Moreover, of these respondents, none express positive sentiment towards the lockdowns and only 2% express neither.

Figure 3.3 shows the sentiment breakdown of those who ID as poor. Overall, we find that 98% of those who ID as poor express negative opinions of the lockdowns.

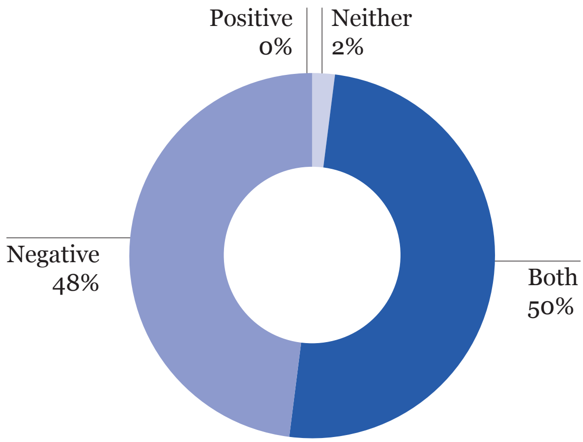

Figure 3.3. Sentiment distribution of those who identify as Poor

3.1.1.3. Economic Impact

References to economic impact focused on one of three areas: food, work, or income. These themes were most often within a context of loss or scarcity. For example: "The income source has definitely been destroyed by the lockdown situation. No work means no income, means no food."

These themes were most prevalent in this dataset, accounting for the majority of respondents. These results reflect the prevailing discourses of the effect of the lockdown on livelihoods.

3.1.2. Theme-Sentiment connections

As we see here with "poor" themes tending to carry negative sentiments, we find that, generally, there is a connection between these themes and the sentiment categories. That is, responses providing positive sentiment, tend to focus on discussions of "safety", while negative responses overwhelmingly focus on themes of poverty or economic impact of the shutdowns.

People expressing that the lockdowns were a positive thing often say so by appealing to a notion of safety or protection such as: "No, this decision, the government did not take a bad decision at all for me. It was a necessary decision to prevent the spreading of the corona virus. The government always wants to protect the mass we should know this." These responses appeal to a common discourse of lockdowns aiding in slowing the spread of COVID-19 and of the health dangers of COVID-19. Overall, 27% of positive responses engage in this type of discourse.

Negative answers, meanwhile, focus on the detrimental effects of the lockdowns which, expectedly, inordinately effect the disempowered and those without means (i.e. "the poor"). Additionally, the outworking of those effects are usually within the context of that poverty, therefore mentions of "work", "food", and "income" are usually in the context of loss or scarcity. Even though these patterns may not be particularly surprising, they are no less important to highlight. Nor are they without interest, especially when we dig into the "Both" category.

As a reminder, "Both" is the sentiment category containing both positive and negative expressions of sentiment. In this dataset, "Both" patterns with "Negative" in that, while there are moments where those respondents mention positive opinions about the lockdowns, they often follow those expressions up with discussions of the detrimental effects in economic terms or the negative impact on impoverished people. “It may be good if it is imposed, but we suffer in all kinds of ways. If lockdown is imposed, it is beneficial for all people, but for us it becomes hard for our livelihood.” Additionally, as seen in this example, the positive parts tend to be very short, unnuanced portions of the response; they lean towards a “generic” positive – a sort of “lip-service” positive. That is, the positive opinions expressed by these responses are not fully adhered to by those respondents. Rather, they express a positive sentiment to fulfill a presumed social obligation towards that sentiment, but then quickly move into more impassioned discussions of negative expressions.

3.1.2.1. Generic Positives

Taking this notion of “generic positives”, we examined further the “Positive” category and found that over three quarters were “generic positive”, consisting of only simple, un-nuanced, non-arguments, such as:

> “No, what the Government is doing, they are doing it for good, and people.”

In contrast, some others provided much more nuanced responses, showing more thought and engagement in their response:

> “Now this lockdown is not bad at all. Still, thousands of people are dying daily despite the imposed lockdown. It would have been a lot worse if it was not for the lockdown. That would have been a great loss for our country Bangladesh, is not that so?”

These responses were generally 50% longer than the generic positive responses, and these respondents tended to speak with raised pitches and intensity, reflecting more engagement (Table 2), interest, and confidence. We termed these responses as "nuanced positive". Figure 3.4 shows the distribution of nuanced and generic positive responses relative to all positive responses.

Table 3.1. Speech Signals of positive responses by type

Positive Response Type	Pitch	Intensity	Pitch Variation	Duration
Generic	0.02	-0.08	0.16	11.77
Nuanced	0.14	0.06	-0.07	28.41

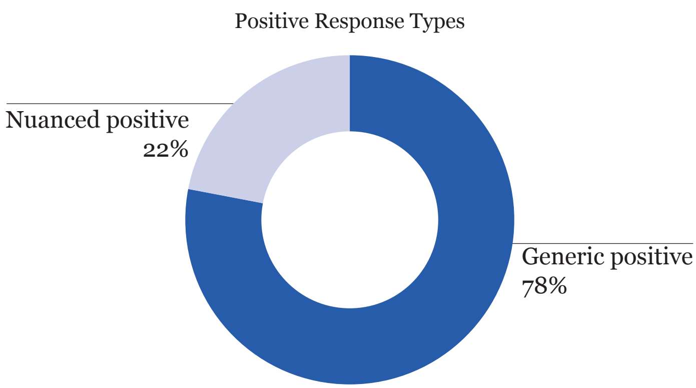

Figure 3.4. Distribution of the two types of positive responses

3.1.2.2. Strong Negatives

In the negative responses, we observe a very different pattern. When we looked at the negative sentiment as well as the "Both" sentiment, we found that there were word choices where people were expressing "strongly negative" sentiment:

...people are dying because of the lockdown.

That is, some respondents answered regarding similar problems (the difficulty in obtaining food, say) but with very different kinds of expressions: "Poor people can't get food"

"We are starving."

There is little doubt that the use of "starving" here, carries a stronger negative connotation and expresses a stronger negative sentiment than "can't get food". Therefore, we conducted another exercise whereby we created a layer of analysis coding for these heavier expressions using an added Topic Classification model. Importantly, this analysis is not on the "Negative" sentiment category alone but crosscuts the sentiment categories. We found that these strong negative expressions tended to come from respondents who identified as poor (Table 3) and, in acoustic analyses, came along with neutral pitch and intensity levels and very low movement of pitch across the responses, indicating very "flat" voices. These speech signals are consistent with a pattern of sadness or resignation.

Table 3.2. Speech signal patterns of Strong Negative responses are driven by those who identify as poor

Strong Negative	Count	Pitch	Intensity	Pitch Variation
Total	62	0.05	0.08	-0.16
Does not identify as poor	15	0.03	0.11	-0.01
Identifies as poor	47	0.06	0.07	-0.21

Within the Negative and Both categories, this "Strong Negative" made up a large proportion of respondents (Figure 3.5).

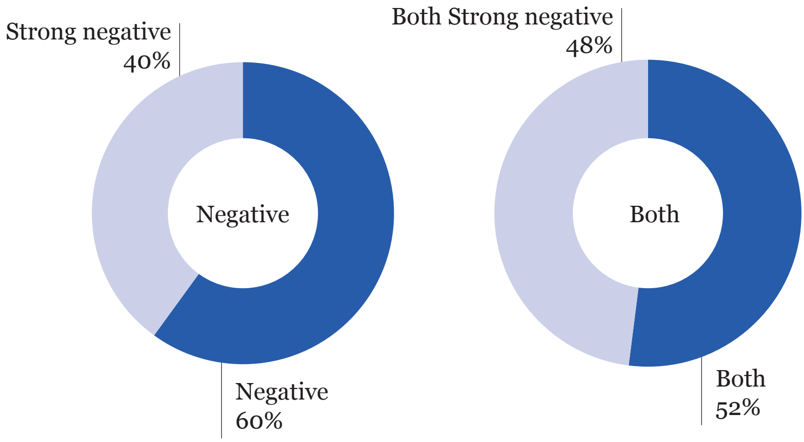

Figure 3.5. Distribution of Strong Negative responses across the original "Negative" and "Both" responses from Figure 3.1

3.1.3. Discussion: Theme Sentiment Connections

These results, taken together, show a very different picture to how the enumerators coded for this question. The results taken from the enumerator coding showed nearly half the sample saying the decision was "good" while the majority of the remainder are coded as "other."

In contrast, the NLP and sociolinguistic approach showed a much more nuanced and less optimistic insight. The only real expressions of "good" – as per the discussion above – were those who were "nuanced positive", which was only 11% of the sample. Those who are neither positive nor negative didn't really answer the question, and generic positive didn't really mean what they said. All together those responses make up 40% of the sample. However, those who were negative are 49% of the sample, and nearly one-third were strongly negative. While the enumerator coding paints a rosy picture of public opinion, this analysis instead paints quite a gloomy one (Diagram 3.1).

Enumerator coding versus NLP + Sociolinguistic coding

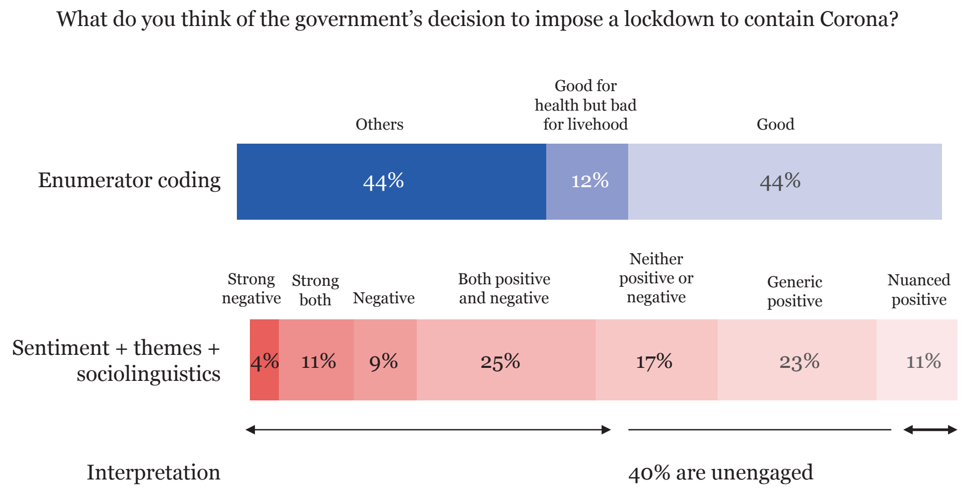

Diagram 3.1. Segmentation of responses to the question "What do you think of the government's decision to impose a lockdown to contain Corona?"

Figure 3.6 provides some detail into why this discrepancy in coding and analysis occurred. Each section shows the Decodis analysis with individual bars representing the three types of enumerator coding. As seen here, Enumerators largely coded "generic positive" as simply "good." This can be given the previously established coding, which could not predict that people in this category would be unenthusiastic or unnuanced at best or feel compelled at worst. They also coded "other" in those categories where respondents either mentioned both or neither. In this way, the enumerator coding both overstates and oversimplifies the “good” position and underrepresents the breadth of nuance in other positions.

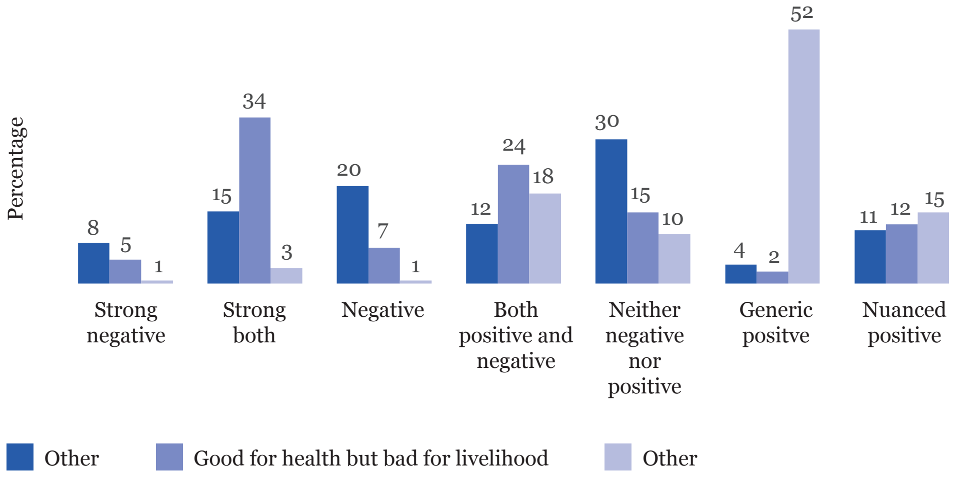

Figure 3.6. Enumerator coding for “What do you think of the government’s decision to impose a lockdown to contain Corona?” within each segment of NLP+sociolinguistic coding

3.2. Requesting Relief

This section examines insight into two facets of relief discussed by respondents: what relief would be most helpful beyond food and what hesitancies do they feel in requesting relief. These data are drawn from over 300 participants across questions 79.9 and 79.10.

3.2.1. Specified Items

Table 3.3 indicates the spread of answers from participants on what kinds of relief would be most beneficial. Here, we see that, the most requested item for relief was “cash” (45%), with “jobs or work” next at 20%. Interestingly, even though the formulation of the question distinctly asks “besides food”, many respondents (16%) respond that “food” is their most needed relief item. This flouting of the questions intent may indicate an urgency in those responses for food relief. It is important to note, as well, that these are not mutually exclusive categories and many respondents gave answers with multiple items (41% of the time).

Table 3.3. Response categories to what kinds of relief would be most beneficial

Specified Items	
Item	% mentioned
Food*	16%
Housing materials	2%
Livestock	5%
Business stock or loans	9%
Need schools to be opened	1%
Cash	45%
Jobs or work	20%
Don’t need (“Can’t wait” for anything)	2%

*10% of the sample specified food items such as rice, lentils, potato and oil.

We further drilled into these results by cross-examining with background questions, demographics, and attitudes. First, we examined these answers by income bracket. Figure 3.7 details the distribution of respondents by income bracket within each kind of answer. Across the income brackets, “cash” is, by far, the most frequent response except in the second highest income bracket (4001-6000 taka per week). For this group, food is mentioned a similar amount as cash. Additionally, we see that respondents above 6000 taka (the highest income bracket) do not provide "food" as a response at all.

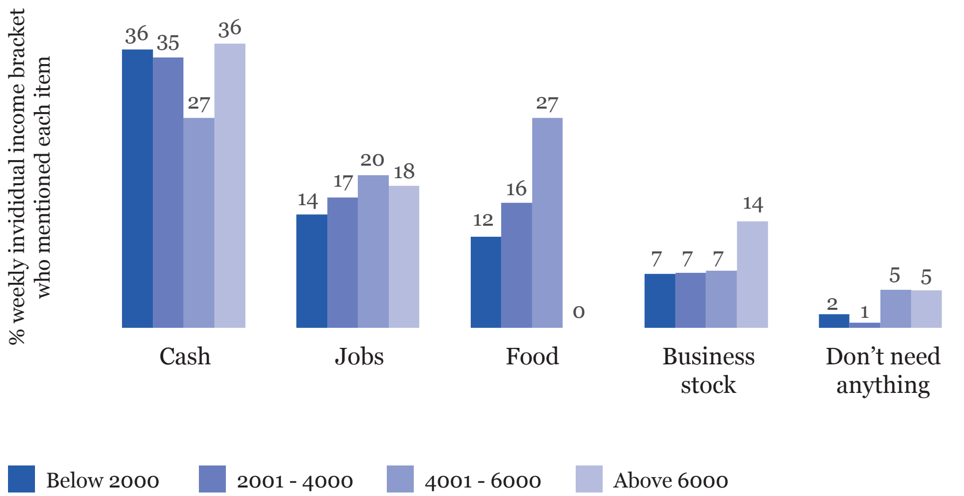

Figure 3.7. Relief item categories by income bracket

Next, we examined various demographic and background information. We found no gendered differences as men and women both mentioned similar items at similar frequencies, with women mentioning cash only slightly more and men being slightly more distributed in their answers (Figure 3.8).

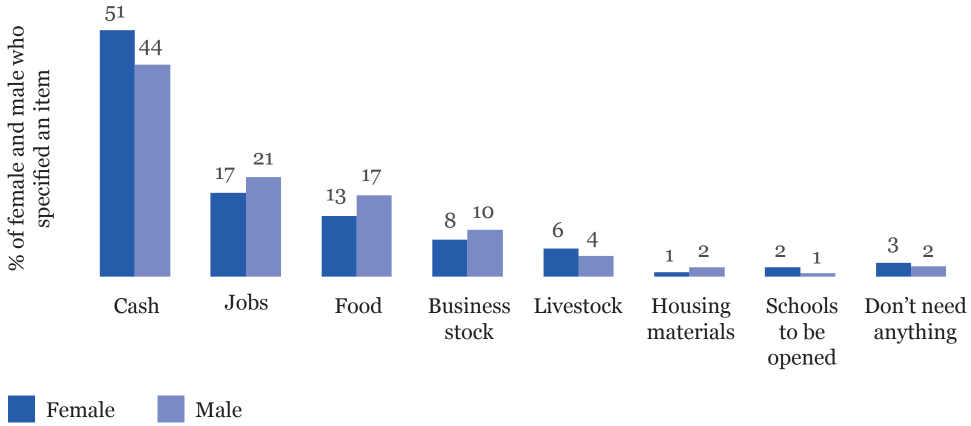

Figure 3.8. Gendered distribution of relief item mentions

We did, however, find difference in response depending on rurality. Urban respondents were more likely to specify items across the options whereas rural respondents were more likely to say they didn't need anything. Additionally, of the specified items, rural respondents requested food the most, but urban respondents requested business stock (Figure 3.9).

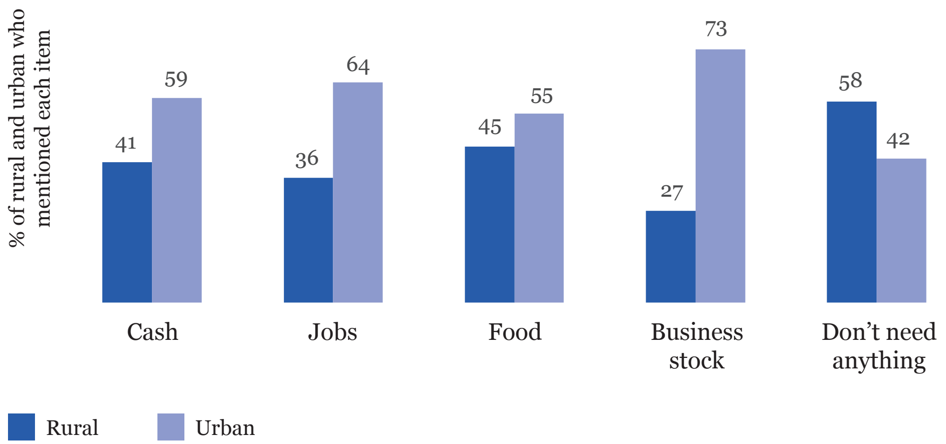

Figure 3.9. Relief item mentions by living condition (rural vs urban)

Finally, we also found that day laborers and salaried workers asked for items across the range, but business owners were far more inclined to ask for business stock (Figure 3.10).

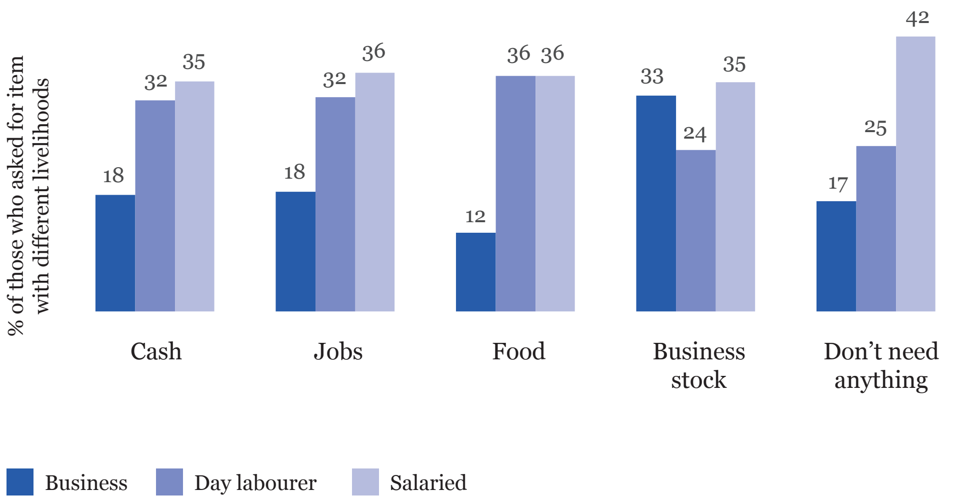

Figure 3.10. Relief item mentions by livelihood

3.2.2. No item specified

Interestingly, while most respondents did detail some sort of item to be helpful, several respondents (21%) specified no item at all. These respondents are mostly from lower income brackets (Figure 3.11) and, crossing their answers with 72.3, expressed a range of sentiments towards the lockdowns (Figure 3.12.).

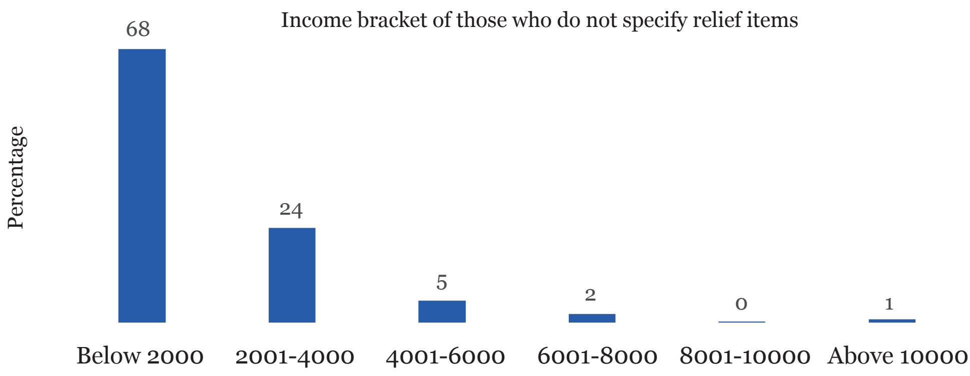

Figure 3.11. Respondents who do not specify items are from the poorest income brackets

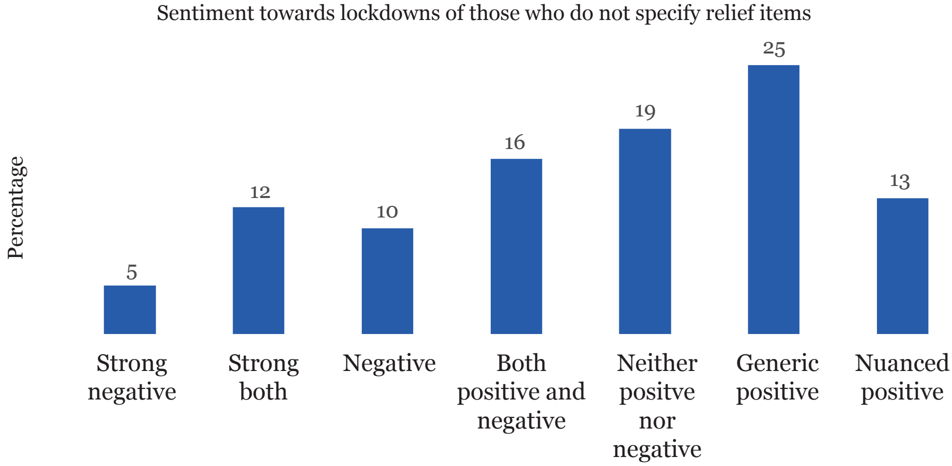

Figure 3.12. Respondents who do not specify items span a range of sentiments towards the lockdowns

With this group being primarily lower income, we may presume that they would most benefit from assistance. However, even though they would benefit the most, they are failing to specify beneficial items. Given that, as seen in the attitudes to the lockdown section, respondents closely associated with being poor (those that ID as poor) expressed their sentiment in a more "resigned" fashion, we may conclude then that these respondents, in answering questions regarding relief, may carry similar sentiments of resignation and therefore not detailing their needs as that would be wasted effort. This may not be resignation to their situation, rather resignation to the likelihood of receiving relief. Indeed, a number of respondents answered in just such a way, saying that they can't sit around and wait for relief: "I work on a daily to daily basis and earn my money and food by this means, how can I sit idly and wait for some people to come and help me and rescue me?"

3.2.3. Hesitancies in Requesting Relief

Beyond specific items, we also asked respondents about whether they felt hesitancy in asking for relief. Table 3.4 details the kind of answers provided by respondents to this question.

Table 3.4. Response types to whether respondents feel hesitancy in asking for help/relief

Hesitancy Responses	
Response	% mentioned
Don't need help	9%
I have no problem in asking for help	11%
Family forbade	1%
Embarrassed (to ask)	34%
Fear of social condemnation	4%
Fear of rejection	3%
Ashamed	8%
Shy	6%
General	19%
Feel bad to ask (because others are in need)	5%

About 1/5th of respondents deny any hesitancy in asking, either explicitly ("I have no problem in asking for help" (11%)) or implicitly ("I don't need help" (9%)). The great majority of respondents, however, do feel they face some sort of barrier to making requests for help.

Respondents who did face barriers in making requests varied in how they expressed these. The way this question was worded in Bangla uses the word shongkot. In addition to "hesitancy", this word also carries a notion of "embarrassment". Over half of respondents reflected this internalized notion in their answer, either through explicitly saying they felt embarrassed to ask for relief (34%) or through a more general affirmative to the question (19%). Still others expressed other kinds of internalized pressures, like feelings of being ashamed (8%) or shy (6%).

Other respondents gave reasons focused on more externalized notions. Several respondents give their hesitancy an externalized force, mentioning a fear of social condemnation (4%) or a fear of rejection (3%). A few respondents also mentioned that their family forbade them to ask for help (1%). Some respondents even discussed feeling bad asking for relief, knowing that other people were in need as well (5%).

3.2.4. Discussion

Table 3.5 illustrates the differences between our coding (NLP Results) and enumerator coding of respondents who mentioned items that would be most helpful or said they didn't need anything. Bear in mind that these results are presented as percent of "mentions", so single responses may contain multiple mentions.

Table 3.5. Comparative coding between NLP and Enumerator Coding

NLP Results			Enumerator Coded Results	
Item	% mentioned		Item	% mentioned
Food	16%			
Housing materials	2%	≈	Home Rent Waiver	0%
Livestock	5%			
Business stock or loans	9%	≈	Loan Assistance	1%
Need schools to be opened	1%	≈	Lower Educational cost of children or provide scholarship	1%
Cash	45%	≈	Financial Assistance	42%
Jobs or work	20%	≈	Arrangement for new income generating activity	11%
Don't need (Can't wait for anything)	2%	≈	Don't need anythin	10%

There is some degree of alignment between the NLP Results and the enumerator results, for instance, "Cash" from the NLP results is in line with "Financial Assistance" from enumerator results, coming in at 45% and 42% of mentions, respectively. However, NLP results provide a much more detailed and nuanced picture.

First, even though the question asks, "Aside from food items…", 16% of respondents mention food and two-thirds of those respondents mentioned specific food items like rice, lentils, potatoes, or oil. This finding highlights just how important food is to the sample. Even though they have been asked to provide different answers, they are taking the opportunity to "drive home" the importance of food. This should not be surprising; as we saw from 72.3, many mentioned the lack of food or the inability to obtain food as a concern, often quite strongly. The need for food and the lack thereof is an important Discourse to respondents. In the enumerator coding, however, there is no mention of food items. The importance of this Discourse is lost in that coding schema.

Relatedly, the need for income and livelihood is highlighted. While money is needed for survival, many respondents are desperate for sustained income, for long-term solutions. These results differ greatly to the enumerator results. Enumerators only coded these sorts of long-term solutions, such as “Arrangement of new income generating activity” or “loan assistance”, a total of 12% of the time. In the NLP results, however, we show that respondents mention “Business Loans” and “Jobs and work” nearly one-third (29%) of the time. Additionally, we highlight “Livestock”, mentioned 5% of the time.

Finally, only a small number of respondents (2%) say that they don’t need anything. This is a drastic reduction from the enumerator coding (10%). The way respondents express this lack of need, too, is of critical importance. Many respondents use expressions indicating that they “can’t wait” for help. They cannot [“sit around and do nothing” quote]. These sorts of expression reveal a lack of confidence in receiving help or relief. Discovering this type of stance is only possible using these kinds of methods.

Of those respondents who did not specify an item, enumerators coded their answers as requesting “financial assistance” 77% of the time, with “arrangement for new income generation activity” taking another 20% (Figure 3.13).

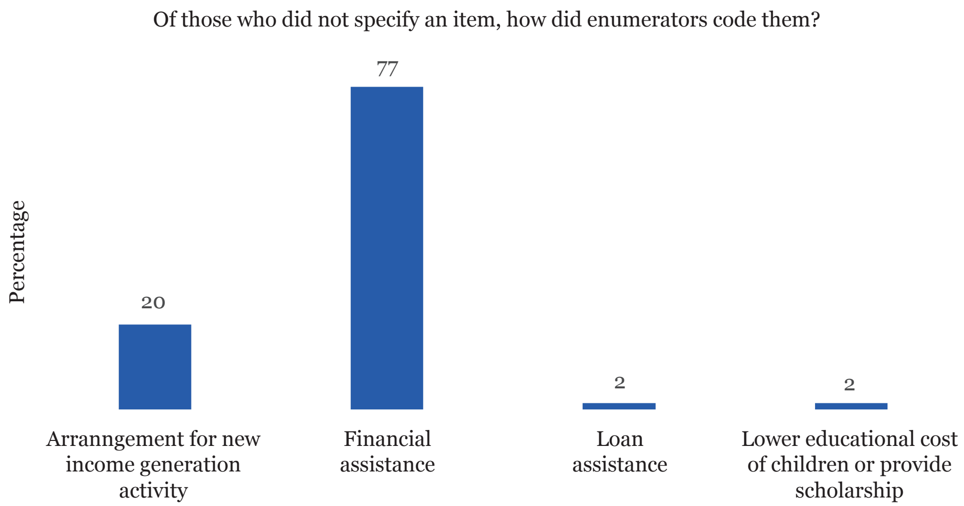

Figure 3.13. Enumerator coding of respondents who did not mention a specific item of relief

While there is some equivalency between the NLP results and the enumerator coding, the results are largely different and paint a very different picture. NLP methods provide a great deal of nuance and insight missed by enumerator coding. Additionally, these results highlight the detriment of "pre-selected" options for broad impact questions versus the strength of NLP methods probing open-ended responses.

In both the NLP analysis and enumerator coding, embarrassment was most commonly noted. As in the previous question, however, NLP results provide a greater deal of insight. In their responses, many answer generally that they are embarrassed to ask. However, we did find there was variation in the sources of that embarrassment. That is, some expressed internalized embarrassment in being "ashamed" or "shy", while others expressed a more externalized force in fear of social condemnation or rejection.

3.3. Enumerator Discourse Analysis

During this project, we also analyzed enumerators for several specific aspects of discourse. Specifically, we measured:

- Talking Amount: The length of time the enumerator spent talking across the entire question segment with the respondent;
- Talking over (Overlap): The amount of time the enumerator spent talking while the respondent was talking;
- Post talk: The amount of time the enumerator spent talking after the respondent began their answer to the question
- Type of post talk: Post talk was classified as either "feedback" (um, yes, etc.), which is unlikely to influence respondent answers, or non-"feedback", which is contentful, reactionary, and potentially influencing of responses

Generally, an increase in any of these measures (talking more, talking over, and non-"feedback" post talk) is undesirable behavior from an enumerator/ interviewer since it may guide, constrain, or otherwise influence the respondent's answer.

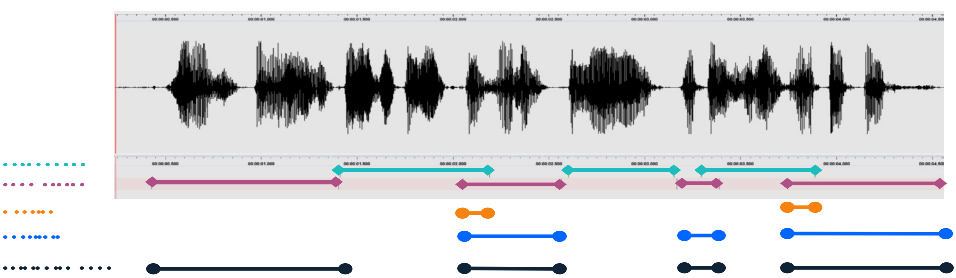

Diagram 3.2. Speech signals and explanations of different types of discourse analysis indicators

3.3.1. Talking Amount

First, we will look at Talking Amount. In Figure 3.13, we show the relative amount of talking enumerators do per question on average. Keep in mind that this represents the percentage of time for the question. That is, a result of 50% would mean that the enumerator and the respondent spoke the same amount of time and had an equal share of the conversation. Anything over 50% means that the enumerator spoke more than the respondent and is, therefore, likely dominating the conversation.

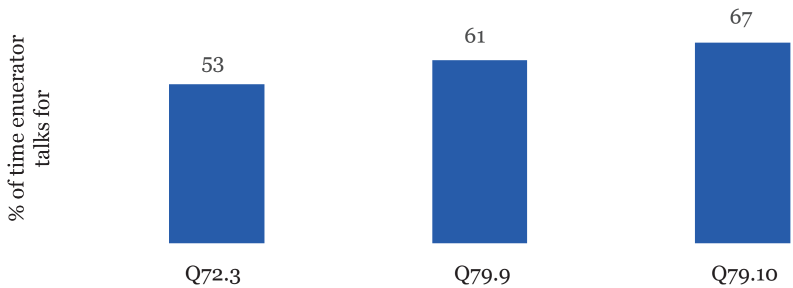

Figure 3.13. Talking amount increases in later questions

Here, we see that enumerator talking amount increases as questions progress. A possible explanation for this is that questions 79.9 and 79.10 are more complex questions and require a greater amount of personal reflection than 72.3. As such, respondents may request more clarification from the enumerators or enumerators may feel a need to prod respondents or ask them to clarify their answers as they may respond in less expected ways.

In Figure 3.14, we see the performance of each enumerator. We see that very few enumerators approach a fully equitable share of the conversations on average.

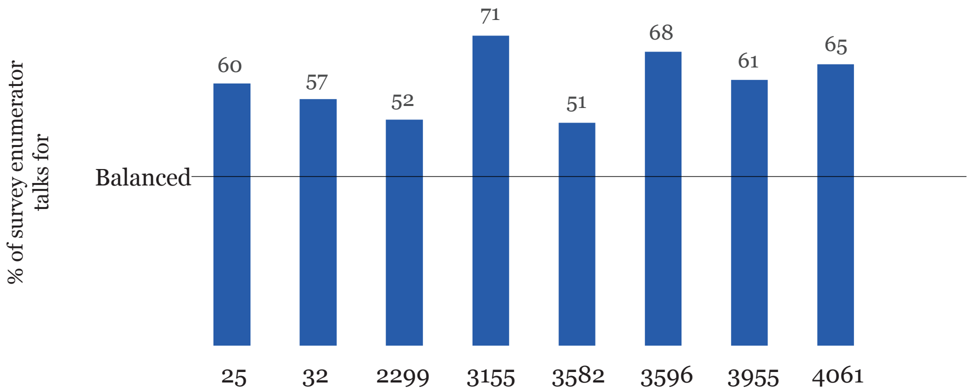

Figure 3.14. Talking amount greater than 50% for all enumerators

While talking amount may be an indicator of who is dominating a conversation, the enumerators alone are not to blame here. Respondents may willingly "give up the floor" or may even push enumerators to talk more, either through requests for clarification or by being less forthcoming in their answers.

3.3.2. Talking Over

If we look at Talking Over (overlap), however, we can investigate how often enumerators are interrupting or failing to give up the floor to respondents. In this way, enumerators assert their power in the conversation and, the more often they do, diminish the respondent's power and, consequently, their willingness to speak.

We examine these overlaps in two ways. First, we look at what percentage of the respondent's answer is talked over by the enumerator. Here, the higher the percentage, the stronger the assertion of power by the enumerator in the conversation. They are either wresting the floor from the respondent or refusing to yield it. Second, we look at how often enumerators interrupted respondents as a percentage of turns interrupted. Here, a turn is defined by when a speaker begins to speak to when they stop talking. This measure represents the frequency of power assertions. Taken together, these two measures provide different facets of power dynamics between enumerators and respondents in these conversations. Figure 3.15 and Figure 3.16 provide the average measures for each enumerator across all three questions for the amount of talking over and the frequency of interruptions, respectively.

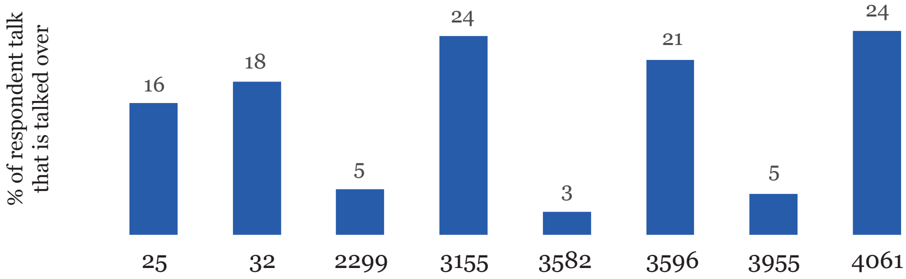

Figure 3.15. Amount of respondent speech that is talked over by enumerators

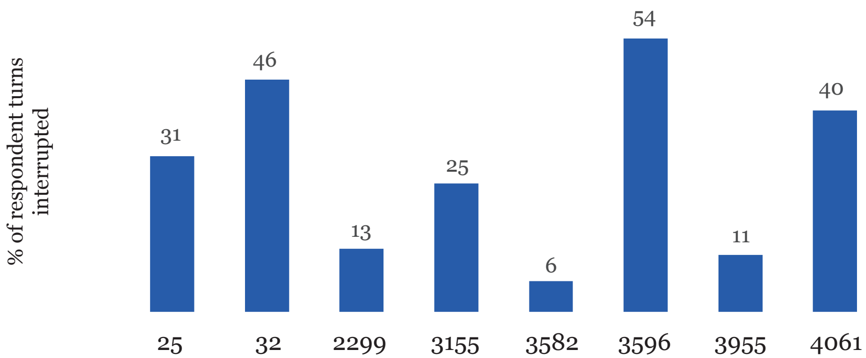

Figure 3.16. Frequency of interruption by enumerators

3.3.3. Post Talk

Finally, we examined Post Talk indicators. Post Talk is measured from the moment the respondent begins speaking after the enumerator has finished asking the question. As such, it provides an indication of potential influences on respondents' answers.

First, Figure 3.17 indicates the average amount of post talk each enumerator engages in across all the questions.

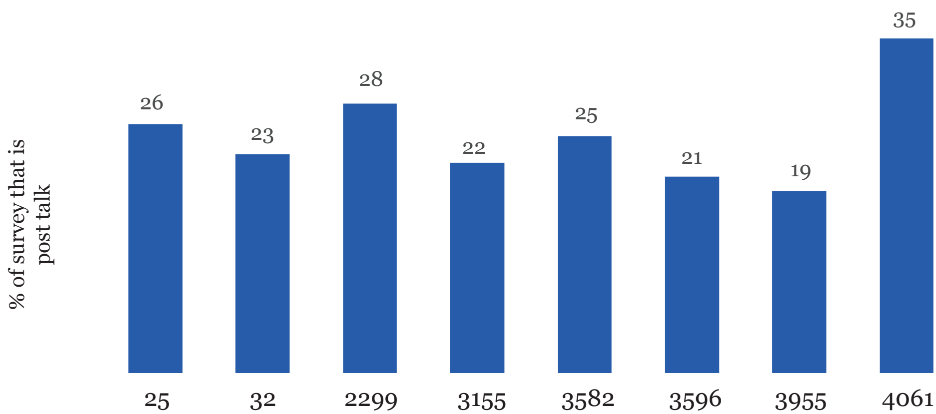

Figure 3.17. Amount of survey enumerator spends talking after the question is asked

Here we see that, on average, a quarter (25%) of the survey consists of the enumerator's talk. Put differently, this number reflects the amount of potential answer-time the enumerator is taking up and the amount of the survey they are taking up after asking the question. Consequently, the higher the percentage in this measure, the greater the risk of the enumerator influencing the respondent's answer.

Second, we also examined the kinds of Post Talk enumerators engaged in. Specifically, in this exercise, feedback such as “yes”, “hmm”, or “sure” (sometimes called “back-channeling”) would be classified as Post Talk. Therefore, we categorized Post Talk moments as either feedback or “content-full”. Feedback is usually used by one conversant to show another conversant that they are listening and paying attention. This sort of Post Talk is not concerning and is, instead, affirming of respondent’s power and encourages answers. “Content-full” talk, however, represents potential leading, priming, confirmation, or counter-argumentation. Content-full talk represents a much greater influence on responses. Figure 3.18 indicates the average percent of post talk by respondent that is content-full and not feedback. These very large numbers illustrate a great deal of potential influence on respondents’ answers.

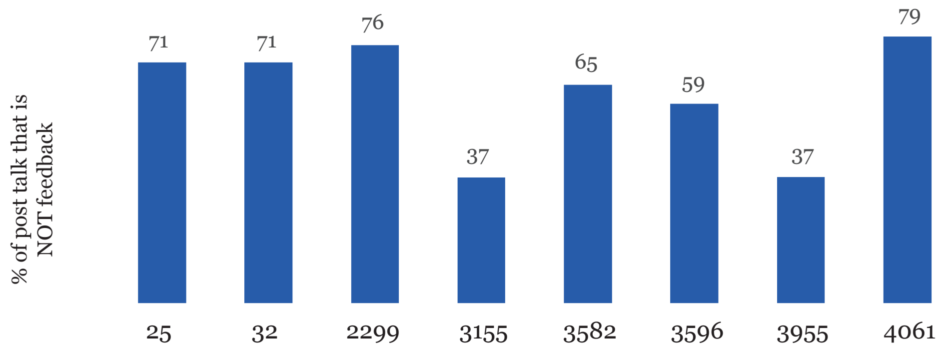

Figure 3.18. Amount of talk after question is asked that is “contentful”

3.3.4. Discussion

Our major point for this section is that each enumerator performs “badly” against these indicators in different ways. In other words, the extent to which enumerators have any of the above behaviors is not systematic –

there aren't just some "bad" enumerators that perform poorly on all of these and some "good" enumerators that perform well. Instead, each of these enumerators performs poorly on one of these indicators. Enumerators, as human interviewers, are subject to the habits and social constraints of conversations.

Our solution is that, ideally, enumerators should simply not be doing interviews, but we should rather use Interactive Voice Recordings (IVR) to do the interviews. However, there are times – such as when a respondent does not have a phone or only has access to a shared phone – where a live interview may be needed. In those situations, these learnings can both help train field researchers and to quality check their interviews if they start falling into these types of bad habits.

4. Conclusions

With this project, we set out to test the extent to which these kinds of mixed-methodologies on open-ended responses on large-scale surveys like this can provide additive value.

We believe that, through these analyses, we have shown the strengths in these methods in discovering quality insights. In 72.3, we displayed means of probing sentiment and how these methods allowed for the contextualization of respondents' answers accordingly and capturing nuance. In 79.9 we showed a broad range of insight that would otherwise have been completely missed or overshadowed by enumerator coding alone. In 79.10, we showed the possibility of including questions that otherwise do not fit into an enumeration paradigm. In our analyses of enumerator discourse behaviors, we show the potential affect enumerator methods can have on results, indicating the benefits of IVR methods for data collection. Additionally, these behaviors could be trained and checked for in live interview situations. Through all these results, we believe we successfully show the breadth of application of these methodologies.

5. References

Athanasiadou, A. (1991). The discourse function of questions. Pragmatics, 1(1), 107-122.

Blommaert, J., & Bulcaen, C. (2000). Critical discourse analysis. Annual review of Anthropology, 447-466.

Cummings, L. (2005). Pragmatics: A multidisciplinary perspective. 1st Ed. Routledge.

De Ruiter, J. P. (Ed.). (2012). Questions: Formal, functional and interactional perspectives (Vol. 12). Cambridge University Press.

Duszak, A. (2003). Us and Others: Social identities across languages, discourses, cultures. John Benjamins Amsterdam.

Edelsky, C. (1981). Who's Got the Floor? Language in Society, 10(3), 383–421. http://www.jstor.org/stable/4167262

El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern recognition, 44(3), 572-587.

Fish, K., Rothermich, K., & Pell, M. D. (2017). The sound of (in)sincerity. Journal of Pragmatics, 121, 147-161.

Feraru, S. M., D. Schuller & B. Schuller. (2015). Cross-language acoustic emotion recognition: An overview and some tendencies. 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 125-131, doi: 10.1109/ACII.2015.7344561.

Hart, C. (2010). Critical Discourse Analysis and Cognitive Science: New perspectives on immigration discourse. United Kingdom: Palgrave Macmillan UK.

Hincks, R. (2004). Processing the prosody of oral presentations. In InSTIL/ICALL Symposium 2004.

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. Speech Communication, 88, 106-126.

Johnstone, B. (2017). Discourse Analysis. John Wiley & Sons.

Laukka, Petri, Patrik Juslin & Roberto Bresin. (2005). A dimensional approach to vocal expression of emotion. Cognition & Emotion, 19:5, 633-653.

Levinson, S. (1979). Activity types and language. Linguistics, 17(5-6), 365-400.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. J. Acoust. Soc. Am. 49, 606–608.

Scherer, Klaus R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, vol. 40, no. 1-2, p. 227-256

Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. Information Systems, 94, 101582.

BRAC Institute of Governance and Development

By influencing policy and practice using rigorous research and by building fit-for-future capabilities through quality educational programs, BIGD aims to promote innovation and improvement in governance and development processes, leading to a just and prosperous society.

brac institute of governance & development

BRAC UNIVERSITY
Inspiring Excellence

bigd.bracu.ac.bd | info@bigd.bracu.ac.bd