computation

MDPI

Article

Word Counting Is Not Enough: A Syntax-Driven Computational Method for Analyzing Participation

María P. Raveau [1,*], Julián I. Goñi [2], Pía Amigo [3] and Claudio Fuentes-Bravo [4]

[1] Faro, Universidad del Desarrollo, Av. Plaza 680, Santiago 7610658, Chile
[2] Science, Technology and Innovation Studies, University of Edinburgh, Old Surgeons' Hall, High School Yards, Edinburgh EH1 1LZ, UK; jvgoni@uc.cl
[3] Departamento de Física, Universidad Técnica Federico Santa María, Av. España 1680, Valparaíso 2390123, Chile; pia.amigo@usm.cl
[4] Facultad de Derecho, Universidad de Chile, Pío Nono 1, Santiago 6640022, Chile; claudiofuentesbravo@gmail.com
* Correspondence: m.raveau@udd.cl

Abstract

Traditional methods of citizen deliberation, such as mini-publics, have been criticized for their limited scope. However, the implementation of massive deliberative processes creates new socio-technical challenges, particularly at the systematization stage. In general, the task of processing citizen participation is typically met with the general tools of textual analysis and Natural Language Processing (NLP). However, processing participation is not just a technical problem, but fundamentally a problem of normative design whose solution must be coherent with the philosophical, ethical, and institutional decisions that the participatory process embodies. In this article, we show how, using basic syntactic rules, we can produce simple and highly explainable reconstructions of public opinions. This methodology was tested on datasets from Chilean participatory processes, specifically dialogs from *We have to talk about Chile* and the 2023 Constituent Process. We show how our strategy enables further information extraction by identifying different syntactic components, providing deeper insights into the answers, while aligning with the goals of citizen participation. Overall, this study presents a compelling alternative for interpreting large volumes of citizen input, particularly in ways that preserve context and advance a more coherent analytical framework aligned with the normative ideals of deliberative democracy.

Keywords: citizen participation; deliberative democracy; NLP; computational linguistics

Academic Editors: Minzhang Zheng and Pedro D. Manrique

Received: 4 August 2026
Revised: 1 September 2026
Accepted: 5 September 2026
Published: 9 September 2026

1. Introduction

The desire to scale up citizen participation creates new socio-technical challenges. One of them is the processing of natural language when the amounts of data and the restricted timelines do not easily allow all relevant actors to go through all the citizens' ideas. Daniel Berliner [1] sets the scene for the challenge clearly in his introductory section:

> Imagine you have convened a townhall meeting (of residents of a town, employees of an organization, or students of a department). Hundreds of people attended, making in total several hundred spoken contributions. Now it is your task to report some actionable information to a decision-maker who, not having attended and lacking the time to read a complete transcript, needs this information in some condensed form. What do you convey?

Of course, some premises underlie the framing of that question. In traditionally set-up deliberative processes, such as deliberative mini-publics (DMPs), it is the norm that processes involve filtering their own outputs (typically recommendations), either through consensus or voting [2], so that they typically come up with a handful of statements and not "several hundred spoken contributions". This kind of process works as both a democratic mirror of society and a deliberative filter of complex information [3]. However, not all participatory processes are deliberative. Moreover, we can imagine a future in which reflexive governance of international affairs would need to articulate multiple deliberative mini-publics along with other opinion-formation mechanisms [4] in a way that does indeed experience the "too many contributions" dilemma. For example, this could be the case in future Global Assemblies.

The large amounts of text that can be generated in participatory processes are usually systematized using automated text processing techniques. However, techniques based on word counts often lack interpretability, while current techniques, such as BERT or LLMs, open up new possibilities, but they are not a solution on their own. These methods still require proper justification, explainability, and experimentation with real-world data. Overall, we find that the socio-technical challenge of data analysis in large-scale participation lies in balancing usability, analytical depth, and traceability at the system level.

In this article, we demonstrate how leveraging basic syntactic rules allows us to produce simple, highly explainable reconstructions of public opinions, aligned with the principles governing citizen participation processes. We then test this method using datasets from large-scale participatory processes in Chile, the "We have to talk about Chile" dialogues in 2020 and the 2023 Constituent Process citizen dialogues, exploring both the potential and limitations of this approach.

Ultimately, our goal is to inspire new avenues for the analysis and visualization of participation data, particularly in ways that preserve context and advance a more coherent analytical framework aligned with the normative ideals of deliberative democracy. The contribution is not the syntactic techniques themselves, which are well established, but the articulation of the conditions under which their use is democratically legitimate and normatively coherent.

1.1. The Goals of Citizen Participation

Following Bächtiger and Parkinson (2019), citizen participation goals can be organized into five categories [5]:

- Epistemic Goals: Identifying the best possible knowledge and solutions to a given social problem through collective intelligence.
- Ethical Goals: Embodying and instantiating the ethical norm of mutual respect.
- Transformative and Clarification Goals: Modifying the preferences and opinions of the general public, or at least clarifying disagreements.
- Legitimacy Goals: Embodying and instantiating the rightful way to source collective decisions, claims of representation, publicity, and substantive requirements about justice, rights, and rationality.
- Emancipatory Goals: Not only achieving the outcomes welcomed by existing systems and dominating actors, but also challenging oppressive structures and giving self-driven expression to marginalized voices.

The capacity to process information is necessary for all of them, but each creates distinct nuances about how and why analysis is carried out. For instance, emphasizing the ethical or legitimacy goals could lead practitioners to emphasize the procedural aspects over the outputs. Adopting the perspective of transformative goals would likely involve looking at opinions before and after deliberation or comparing them with the general population.

However, among these goals, epistemic and emancipatory objectives carry particular implications for any analysis strategy. Epistemic goals suggest the need to show citizens' ideas with sufficient nuance that they can be understood in their complete structure, not only in their frequency of appearance. Emancipatory goals add a requirement that has rarely been discussed in the specialized literature: the analysis cannot incorporate informational relevance criteria that implicitly determine which voices deserve to be processed, because in a democratic participatory process, all voices carry the same political weight regardless of their linguistic sophistication.

1.2. The Challenge of Systematization

As we said previously, citizen participation processes can result in a large number of contributions, which then need to be systematized. The technical procedures used to "condense information" have to meet external criteria in addition to their capacity to reduce complexity. They are typically expected to be accountable, transparent, and traceable, among other political values. Organizers may not have well-thought-out plans for how the data will be processed [6]. But the scale and timeline of the challenge, as well as the (impossible) desire for knowledge devoid of opinion [7], make computational social analysis the likely candidate.

In that sense, the specific challenge of processing citizen participation is typically met with the general tools of textual analysis and Natural Language Processing (NLP). There is a path dependency from previous text analysis tasks, but also adaptation in use, as well as a mutual shaping of "new" requirements and existing technology [8]. The review by Romberg and Escher [9] organizes the array of machine learning techniques used for citizen participation into three groups: categorization and topic extraction, argument mining, and sentiment analysis. But there is an even broader tool set available for practitioners, from simple word counts (and Wordclouds), semantic networks, and dictionary analysis, among others.

Recent developments in text analysis make use of neural networks to understand the relationships among words in a given text. This is the case for word embedding models, such as Word2Vec [10] or GloVe [11], which give one vector per word, and the distance between two vectors relates to the semantic similarity between the corresponding words. More advanced embedding models like BERT (Bidirectional Encoder Representation of Transformers) [12] are built on Transformers, a deep learning architecture based on the multi-head attention mechanism [13]. However, as these models draw upon deep learning, they bring to the table classical opacity problems related to unsupervised AI [14]. Put simply, what these approaches gain in computational capacity comes at the cost of explainability and simplicity, which are particularly fundamental when it comes to democratic engagement.

Large Language Models (LLMs), on the other hand, have considerably expanded the technical capabilities available. It is important to note that LLMs are technically capable of performing dependency grammar analysis with high precision. However, what these approaches gain in computational capacity they trade for traceability and explainability, which are constitutive properties of an analysis coherent with the democratic process that originates it.

1.3. Algorithm Transparency and the Problem of Epistemic Exclusion

Current algorithms for systematization face some problems with respect to participation goals, which we have already outlined in the previous section. The first is that of algorithmic transparency, which is addressed in the international and national regulations on the use of artificial intelligence systems in contexts of citizen participation and demo-

cratic governance. These regulations have been developed especially since the rise of AI, and they demonstrate that automated processing of citizen participation is a problem of normative design before it is a problem of technical efficiency.

At the international level, the Recommendation on the Ethics of Artificial Intelligence adopted by the UNESCO General Conference in 2021—the first global normative instrument in the field, adopted by 193 member states—establishes that AI systems used in democratic governance contexts must satisfy requirements of transparency, explainability, and auditability that allow citizens and institutions to understand and question decisions produced or informed by those systems [15]. The Recommendation makes explicit that these requirements are conditions of democratic legitimacy: an AI system that cannot account for its results in terms comprehensible to those affected by them is incompatible with the principles of participation, accountability, and non-discrimination that democracy requires. In the same vein, the OECD Principles on Artificial Intelligence establish transparency and explainability as the governing principles of AI use in the public sector, together with effective human oversight and accountability throughout the entire process [16].

At the national level, Chile has translated these international commitments into normative instruments of direct application. The National Artificial Intelligence Policy establishes transparency and explainability in system operation, the prevention of biases and discrimination, and accountability throughout the lifecycle of the AI system, as explicit requirements for the use of AI in the public sector [17]. The Artificial Intelligence Bill currently under consideration in the Senate (Bulletin N°16821-19), explicitly enshrines the principle of explainability, according to which AI systems "will be created, developed, innovated, implemented, and used in a manner such that their results or outputs are comprehensible and intelligible to the people they impact," together with the principles of traceability and human control over automated systems [18].

The second problem has remained systematically undertheorized in the literature on participation processing: the markedly asymmetric distribution of syntactic complexity of individual opinions. In the dataset from "We have to talk about Chile", the median number of words is 19. This brevity is a structural property of heterogeneous citizen participation that reflects the real distribution of linguistic and argumentative capital in the deliberating population [19]. This distribution generates a serious normative problem facing the automated processing of citizen participation: that the informational optimization algorithms prioritize the more complex or informative opinions. A dimensionality reduction algorithm, a frequency-weighted topic model, or a large-scale language model instructed to identify the most informative cases will naturally converge toward that subset, because informational optimization and computational efficiency point in the same direction: process less to obtain more. Large language models are particularly susceptible to this form of bias because their attention architecture weights the relevance of tokens as a function of their co-occurrence with other tokens in the training corpus, which contains predominantly text produced by speakers with high linguistic capital [20].

In the context of democratic citizen participation, this logic is a form of systematic epistemic exclusion. The participant who wrote 'education must be free' contributed to the process no less than the person who developed a twelve-clause opinion: they contributed differently, with the linguistic and argumentative resources available at that moment and in that context. The legitimacy of their opinion does not derive from its syntactic complexity or from its marginal informational value: it derives from their condition as a citizen participant in an institutional process designed to include them. Excluding them from the analysis because their opinion is informationally dispensable systematically reproduces the cultural capital inequalities that deliberative design attempts to correct [21,22]. Participants with greater familiarity with formal political discourse, greater argumentative articulation

capacity, and greater exposure to institutional contexts produce the most complex opinions. When the analysis prioritizes those opinions for informational efficiency reasons, it amplifies exactly the voices that already have greater representation in the public space and silences those that the participatory process intended to incorporate.

1.4. Between WordClouds and WordTrees

We have already seen the problems of algorithmic transparency and auditability that come with the cutting-edge methods, and the epistemic exclusion problems that can result from LLMs or frequency-weighted topic models. But are simpler methods capable of meeting the epistemic requirements of systematizing a participatory process? Below is an example taken from the initiative *We have to talk about Chile*, specifically the question "What should we change, improve, or maintain in Chile?" [23]. Figure 1 shows a Wordcloud of the 100 most frequent words in these answers. No preprocessing was performed, but eliminating stop-words (stop-words are functional words, such as "the", "and", "or", etc., which are often removed in text analysis as they carry little semantic value).

Figure 1. Example of a Wordcloud. The text was originally in Spanish, and no preprocessing was performed other than eliminating stop-words. The Wordcloud was subsequently translated to English. Repeated words are due to the translation.

What is the idea that is being asserted? Could you actually extract ideas from single words? It will all depend on what you consider an idea; perhaps you take away that this particular concept was highly popular. But it will certainly not produce an idea that we can reasonably think of, by epistemic standards, as a potential solution to complex issues.

This problem is not unique to Wordclouds. It represents a broader problem with statistical procedures that consider the "word" as their fundamental unit of meaning. These categorical word representation models [24] can range from simple frequencies (such as Wordclouds) to more complex weighted frequencies of words across documents and corpora, such as Term Frequency-Inverse Document Frequency (TF-IDF), which underpins techniques such as Topic Modeling. Moreover, the general problem raised here—that aggregate counts alone can be insufficient for meaningful interpretation—extends beyond Wordclouds and citizen-participation text analysis techniques [25].

Focusing solely on individual words causes you to overlook the importance of syntax in language [24]. In other words, meaning arises from how words are specifically arranged

within a sentence. A sentence is not just a collection of words; it is a complex structure in which terms modify and influence each other. Techniques centered on "words" often incorrectly assume term independence—that the occurrence of a word is independent of the occurrence of other words [26]. This is obviously not the case when words are semantically related, like "public engagement" or "deliberative democracy". These words form chains of meaning—what we call n-grams (in this case, they are bigrams, as they are a chain of two terms). Bigram networks are visualization tools that represent these semantic relations. Here, nodes are words, and a link between two nodes is established when they form a bigram, where the weight of the link is proportional to the bigram's frequency of occurrence. Although they are becoming more common [27–29], bigram networks are difficult to interpret when there are many nodes, and are less popular than, for example, Wordclouds.

Alternative visualizations have sought to overcome the limitations of Wordclouds and bigram networks. For example, a Wordtree is a graphical version of the "keyword-in-context" method, an analysis that looks at a word (keyword) and each of its appearances in a text to learn its meaning through its context [30]. A Wordtree depicts multiple parallel sequences of words in a tree structure. However, a readable Wordtree usually requires selecting a sample of sentences and a great deal of pre-processing. Furthermore, if a Wordcloud gives a clear idea of frequencies but no context, a Wordtree features the inverse problem: there is a lot to read, but little information on frequencies. Figure 2 shows an excerpt of a (translated) Wordtree from the *We have to talk about Chile* dataset.

Figure 2. Wordtree for the topic "health", starting from the word "improve". For this visualization, a random set of phrases was selected and exhaustive preprocessing was carried out. This consists of reordering and cutting phrases to generate the tree structure.

2. Data and Methods

2.1. Lexical–Syntactic Approach

Our approach is based on the dependency relations among words in a sentence. Unlike what bag-of-words methodologies assume, words within a sentence are structured hierarchically. The dependency structure consists of directed binary relations of "head" and "dependent" between words. Each word (except for the root of the sentence) has a head

word and may or may not have one or more dependents. Each word also has a grammatical role and a part-of-speech tag. Together, all these elements help us organize the information in a meaningful way.

2.1.1. Dependency Relations

The head-dependent relationship is the basic unit of sentence structure, where "heads" are higher in hierarchy than their "dependents". A given word is the "head" of the words that are immediately dependent on it, and those words are its "dependents" and serve as modifiers. The "root" word is the head of the entire structure and does not have a head of its own [31].

2.1.2. Grammatical Role

A head-dependent pair of words can also be characterized by the grammatical function that the dependent plays with regard to its head [31]. This includes familiar notions such as "subject", "object", and "complement", but goes far beyond these. The complete list of grammatical roles is provided by the Universal Dependency (UD) project and can be found in Table 3 of De Marneffe et al. [32]. Here, we focus on the core arguments of a predicate, which essentially refers to subjects and objects.

2.1.3. Part of Speech

In addition to binary relations, each word can be classified according to its individual behavior within the language system. These categories are commonly referred to as "word class" or "Part of Speech" (PoS). Traditional classification includes noun, verb, adjective, adverb, pronoun, preposition, conjunction, and interjection. Modern linguistics has incorporated finer distinctions, recognizing symbols and punctuation [32].

Table 1 shows a set of common grammatical roles and parts of speech, from De Marneffe et al. [32]. Figure 3 shows an example of dependency relations, grammatical roles, and part of speech for the sentence "The State must regulate and guarantee human rights". The encircled label below each word indicates the part of speech (PoS). Relations between pairs of words are displayed with directed arcs from head to dependents, and the tag near the arc denotes the grammatical role of the dependent in the head-dependent relation. For example, take the words "rights" and "human". The arc goes from "rights" to "human", meaning that "rights" is the head of "human". The tag above the arc shows "amod", which indicates that "human" is an adjective modifier of "rights". Note that the verb "regulate" is the root of the sentence, and thus it is the only word that has no head.

Table 1. Common grammatical roles and part-of-speech labels, from https://universaldependencies.org/, accessed on 1 August 2026.

Part of Speech (PoS)		**Grammatical Role**	
NOUN	Noun	nsubj	Nominal subject
ADJ	Adjective	obj	Nominal object
VERB	Verb	obl	Oblique nominal
AUX	Auxiliary	nmod	Nominal modifier
DET	Determiner	amod	Adjective modifier
CCONJ	Coordinating conjunction	appos	Appositional modifier
ADV	Adverb	cc	Coordinating conjunction
ADP	Adposition	conj	Conjunct
PRON	Pronoun	aux	Auxiliary
PROPN	Proper noun	det	Determiners
PUNCT	Punctuation	punct	Punctuation marks

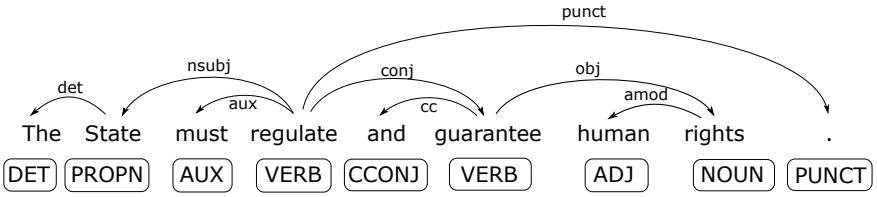

Figure 3. Syntactic elements of an example sentence.

2.2. Syntagmatic Decomposition

The algorithm to extract verb, core argument nominals and nominal modifiers is the most challenging element of the strategy. The result of the algorithm is a component decomposition in the fields "verb", "subject", "mod subj" (modifiers of nominal subject), "object" and "mod obj" (modifiers of nominal objects). The steps are as follows:

1. Identify the root of the sentence. The root is usually a verb, unless there is no verb in the sentence or a core argument nominal is linked to a copulative verb.
2. Identify a nominal subject (nsubj) with a direct dependence on the root. If this exists, it is displayed in the "subject" field; then nominal modifiers (nmod, amod, appos, acl, compound) with dependence on the subject are sought. If there are modifiers, they will be displayed in the "mod subj" field.
3. Identify the nominal object (obj). If the root is a verb, the nominal object may depend directly on it or on another verb that depends on the root. For example, in the clause "I think that the issue of pensions was not addressed.", the root is the verb "think" and "addressed" is a clausal complement dependent on "think". In this case, the nominal object "issue" depends on "addressed", and not on "think", the root verb. Once the final verb in the chain is identified, it is displayed in the "verb" field.
4. Search for the nominal or oblique object (obj, obl) with direct dependence on this verb, and display it in the "object" field. Then seek nominal modifiers with dependence on the object. If there are modifiers, they will be displayed in the "mod obj" field of the component decomposition.
5. If the root of the sentence is not a verb, it is assigned to a nominal object and displayed in the "object" field. The algorithm then searches for a copulative verb with dependence on the root/object and for modifiers of the root/object.

To show the results of the algorithm for extracting verbs, nominals, and modifiers, we used a small sample of 15 sentences (see Table 2). These sentences were chosen to cover a broad range of syntactic structures, including the use of copulative verbs and the absence of verbs. The results of the component decomposition algorithm in this data set are shown in Table 3. As our main purpose is to use this method in Chilean participatory processes, we designed the algorithm to work in the Spanish language, using the Spanish version of Stanza (The specific grammatical roles may change from one version of Stanza to another. Thus, the names we use here, such as "nsubj", "obj" or "obl", among others, are valid for the package version 1.10.1.) for Python (version 1.10.1, [33]). For the purposes of this article, we selected a subset of sentences that worked similarly, using our algorithm, both in Spanish and English. The syntactic decomposition shown in Table 3 is the result of applying our algorithm to the sentences in Table 2, translated from Spanish.

Table 2. Sample of sentences to illustrate the method.

1	4) Promote regional councils
2	We must re-implement civic education
3	Justice must be changed, because it sees nothing, it is asleep
4	If justice were carried out as it says, many people would be afraid of doing what they do
5	We know the needs of the people and we should be able to give our opinion and resolve
6	Immigration regulation because many people have entered who do not come to work and who spread violence and drug addiction.
7	The country must be plurinational and intercultural, in addition to respecting the rights of indigenous peoples.
8	The State must guarantee the care and protection of human rights, avoiding violations of them and compensating the victims.
9	It would be good to have a law on political parties that allows the existence of regional parties.
10	The State must guarantee social rights such as health, education, housing, a healthy environment, and a dignified old age.
11	The issue of immigration is serious in relation to the security that is experienced in Chile.
12	The production paradigm is currently obsolete, considering that it does not protect the preservation of the environment.
13	The lifelong salary should be eliminated.
14	The State has to change education and that a system without gaps between the public and the private
15	The right to gender diversity.

Table 3. Result of the component decomposition algorithm on the small dataset.

	Subj	**Mod Suj**	**Verb**	**Nom**	**Conj Obj**	**Mod Obj**	**Complement**
1			promote	council		regional	
2	we		re-implement	education		civic	
3	justice		change				, because it sees nothing, it is asleep
4	person	many	be	afraid			of doing what they do
5	we		know	need		person	and we should be able to give our opinion and resolve
6				regulation		immigration	because many people have entered who do not come to work and who spread violence and drug addiction.
7	country		be	plurinational	intercultural		and intercultural, in addition to respecting the rights of indigenous peoples.
8	State		guarantee	care		right human	, avoiding violations of them and compensating the victims.
9			have	law		party political	that allows the existence of regional parties.
10	State		guarantee	right		social health	, education, housing, a healthy environment, and a dignified old age.
11	issue	immigration	be	serious			in relation to the security that is experienced in Chile.
12	paradigm	production	be	obsolete			, considering that it does not protect the preservation of the environment.
13	salary	lifelong	eliminate				.
14	State		change	education			and that a system without gaps between the public and the private
15				right		diversity gender	

2.3. The Strategy

The strategy consists of extracting syntactic components: actions (verbs), core arguments associated with actions (subject and object), and their modifiers. Then, we select the components to visualize, along with a set of terms for each component, and some random sentences as examples. These elements are combined into a visualization of predicative components, where font size is proportional to term frequency.

0. Pre-processing and syntagmatic decomposition

The first steps of pre-processing depend on the nature of the data. If the text comes from a dialogue transcription, it usually includes the participant's name followed by their speech. Since the analysis is carried out at the sentence level, each sentence must be examined separately. Thus, preprocessing will include deletion of the speaker's name and sentence tokenization (for long sentences, an additional step of independent clause separation may be convenient). Common text pre-processing techniques—such as lowercase, punctuation and stop-word removal—are neither necessary nor recommended at this stage. Unlike other NLP methods where feature dimensionality reduction is convenient, the tools for linguistic analysis we use here were trained to understand real human language. After preprocessing, each sentence is fed to the syntagmatic decomposition algorithm to extract verbs, core argument nominals, and nominal modifiers.

1. Choose the main component

Since our method is based on the identification of sentence components, the head of the strategy is the selection of the main component. This selection depends on the nature of the question/instruction that motivates the dialogue. Three types of questions can be distinguished: those that raise topics (suggesting extraction of nominal core arguments); those that elicit actions (suggesting extraction of verbs); and those that elicit evaluative statements (suggesting extraction of adjectives).

It often happens that the answers include a mixture of the previous types. Consider the following question (*Consultas Ciudadanas* [*Citizen Consultations*], 2020): "What does a good neighborhood mean to you?" In this case, many answers were formulated with adjectives, such as "clean", "safe", and "quiet". But it was also common to mention things that should or should not exist in a good neighborhood, such as "green areas" or "crime". In this case, separate analysis can be performed with adjectives and nouns, respectively. It may also be the case that we are interested in some specific piece of information from the answers, regardless of the type of component that the question elicits. For example, in *We have to talk about Chile*, one of the questions was "What should we do to achieve this?". This question elicits actions, and thus, a verb extraction would have been the logical choice. However, we could be more interested in the private or political actors mentioned in the answers, thus extracting nominal core arguments instead of verbs.

2. Select a set of terms

Once the main syntactic component has been identified, it must be extracted from all sentences in the corpus to produce an initial count. This count will generate a list of terms with their respective frequencies. The list must be refined to eliminate irrelevant terms, such as semi-auxiliary verbs or overly general words. Terms that are part of the question itself or that are too common due to the dialogue context should also be removed. If the number of terms is too large to handle, it is recommended to select a subset of relevant terms, by frequency or by importance given the project's goal. Another option, if the list contains semantically similar terms, is to group them into categories.

3. Choose additional components to visualize

Additional syntactic components can be selected according to their dependency relationship to the main component. For example, if the main component is a verb, proper candidates for additional components would be nominal core arguments—such as nominal subjects and direct objects—as they have a direct dependency on verbs. If the main components were, in turn, nominal core arguments, the natural candidates would be nominal dependents, like nominal or adjectival modifiers. Afterward, additional components must be extracted for each term in step (2), also selecting a subset of relevant terms.

4. Build a visualization of predicative components

The final step is to visualize the selected components and terms in an ordered way, using a matrix-based structure where columns are components and rows show terms and their frequency. Font size is a function of term frequency. A randomly selected example sentence illustrating the main component, the nominal, and one of its modifiers is included for each entry.

At each of these steps, methodological decisions often have to be made, such as discarding uninformative terms or grouping terms into categories. The selection of grammatical components is aligned with what the question elicits and the information one seeks to extract. These decisions aim to produce a more informative systematization and are present at all stages of the strategy, except in the syntagmatic decomposition. All of this introduces subjectivity into the analysis. Therefore, to ensure the replicability of the analyses, it is essential that the systematization report document all decisions made by the team of analysts. This makes it possible not only to replicate the results, but also to interpret them in light of the decisions made.

2.4. Data

We have to talk about Chile, a massive dialogue exercise

We have to talk about Chile (TQH, by its Spanish initials) was a project organized by two of Chile's most traditional universities, after massive political unrest in October 2019. It consisted of a series of digital conversations in which more than 8800 people discussed the future of the country [23]. The project objectives were (i) to promote a massive social conversation about the challenges of the country, (ii) to promote a way of dialogue that values our differences, and (iii) to rigorously systematize the vision of the future of Chilean society [23].

The dialogue structure consisted of four questions: 1. What has been the predominant emotion of the past week? 2. What do we have to change, improve, or keep in Chile? After all participants answered these questions, the group chose a topic to explore in depth. The following two questions are related to the selected topic: 3. In relation to the prioritized topic, how can we achieve that goal? 4. What can I do as a citizen to achieve it? Each dialogue was guided by a facilitator who also registered each participant's intervention. Facilitators were trained to annotate complete syntactic sentences using the SPOCA structure (subject, predicate, object, complement, adjunct), which made further systematization easier. This validation step can be characterized as a form of member checking [34], a technique used in qualitative research to strengthen the credibility of participant-reported data by returning the researcher's rendering of a statement to the participant for confirmation. Importantly, this validation took place in front of the deliberating group rather than in a private exchange between facilitator and speaker, which distinguishes it from dyadic respondent validation.

The final database consists of 99,630 rows, one row per participant intervention. Of these, 38,606 rows answer the question "What do we have to change, improve, or keep in Chile?", which is the question we will focus on later. The average length of the answers was 19 words.

Citizen Dialogues, Chilean Constituent Processes 2023

The 2023 Constituent Process comprises different citizen participation mechanisms: citizen consultations, popular norm initiatives, public hearings, and citizen dialogues [35]. In self-convened citizen dialogues, participants were organized in groups of 4 to 6 people to discuss one of the 11 topics of constitutional interest. These topics were previously selected and validated by experts from the Secretariat of Citizen Participation due to their relevance to the constitutional draft. After selecting a topic, the participants were invited to answer a question on the topic and also to justify why they chose it and what aspects of the constitutional draft caught their attention.

Each question was answered in an open response format and digitally recorded on a web platform. A total of 10,143 people participated in 2267 dialogs. The systematization was carried out by an interdisciplinary team from different Chilean universities. Unlike TQH, participants self-recorded their responses without a standardized format, which introduces greater variability in the syntactic properties of the corpus and poses greater challenges for automated processing. This contrast between the two databases is itself analytically informative: it shows how the facilitation protocol directly affects the quality and regularity of the corpus available for lexical–syntactic analysis.

The final database consists of 2267 rows, one row per dialogue. Each group chose one of the 11 questions to discuss. We selected the answers to the question "What should be the fundamental rights and freedoms of individuals?", the second most frequently chosen question, restricting our database to 440 dialogs. The average length of the answers is 34 words.

3. Results

In this section, we present the results of applying the strategy outlined in Section 2.3 to the two databases described in Section 2.4. These databases were chosen because of the distinct analytical challenges they pose. The *We have to talk about Chile* database is extensive, and automatic extraction of information is crucial. However, the presence of facilitators during the data collection process aids in the structured segmentation of syntagms. In contrast, the *Chilean Constituent Process* (2023) database is less structured, as participants self-recorded the dialogs without a standardized format, adding significant challenges for pre-processing and making the analysis more complex.

3.1. We Have to Talk About Chile

3.1.1. Pre-Processing and Syntagmatic Decomposition

For this database, we focus on the question "What do we have to change, improve, or maintain in Chile?". First, for all answers with more than one sentence, we performed sentence tokenization using the Stanza module in Python. Data collected with the assistance of facilitators resulted in sentences that primarily adhere to a grammatical protocol, which facilitates the syntagmatic decomposition of the algorithm. However, not all sentences conform to this standardized structure. In many cases, sentences exhibit greater complexity, unclear structures, and multiple embedded ideas. To address this, an algorithm was developed that separates sentences based on the identification of independent clauses. After clause separation, the syntagmatic decomposition was applied to a total of 49,819 independent clauses.

3.1.2. Main Component

In the question "What do we have to change, improve, or keep in Chile?", the most effective approach is to focus on sentences containing the three main verbs in the question—change, improve, and maintain—and to extract the nominals associated with

them. The results indicate that "change" and "improve" are the predominant verbs, alongside auxiliary verbs such as be, have, and do.

3.1.3. Additional Components

The most frequent nominals associated with the verb "change" clearly indicate the primary aspects that respondents wish to change in Chile: "education", "constitution", "health", and—with the highest frequency—"system". The nominal "system" is somewhat ambiguous and can carry multiple meanings. Analyzing the modifiers associated with it reveals that the top three are "pension", "health", and "education". Interestingly, while "health" emerges as the second aspect to change within the nominal "system", it ranks fourth in the nominals associated with "change." This illustrates the capacity of the strategy to uncover deeper layers of information embedded within sentences. The same applies to the term "judicial." Although it does not appear among the top five nominals for any verb, it does rank in the top three for the nominal "system" under the verb "improve."

3.1.4. Visualization

Figure 4 summarizes the results of the steps described above. The table presents the components hierarchically, starting from the main component selected in this analysis: the three main verbs "change", "improve", and "keep". As with a Wordcloud, the font size represents frequency ranking. However, the table also includes the exact frequency in parentheses, providing additional information to the reader. Each main verb is followed by its associated nominal objects, with a list of modifiers specified for each nominal. Finally, the table provides an example sentence, randomly extracted from the data, containing the main component, the nominal, and one of its modifiers. This visualization is a direct answer to the central question posed in the analysis of the data: "What do we have to change, improve, or keep in Chile".

The methodological decisions underlying the previous results are: (i) "constitution" and "Constitution" are treated as equivalent, and their objects and modifiers are combined; (ii) the nominal "country" within the verb "maintain" was removed, because it is assumed to be an indirect object; (iii) the verb count includes all instances, whether or not an associated nominal is present. For the analysis of nominals, by contrast, empty fields are not considered.

3.2. Citizen Dialogues, 2023

3.2.1. Pre-Processing and Syntagmatic Decomposition

From this database, we chose the question "What should be the fundamental rights and freedoms of individuals?", with 440 responses. Responses were self-recorded on a webpage without a standardized format, resulting in varied punctuation and writing styles. Many responses were written as lists of ideas rather than paragraphs, enumerated using numbers or symbols, with line breaks that may or may not be accompanied by periods. Since this variability poses difficulties to dependency parsers, a crucial preprocessing step involves cleaning these symbols while preserving essential punctuation. After text cleaning, the same clause separation method was applied, obtaining a total of 1428 clauses.

change (7338)	system (1211)	pension (172), health (147), education (57)	Change the pension system because it is not enough to live on.
	constitution (703)	new (8), politic (8), current (4)	We citizens need to change to a new constitution, making a profound change in politics and moving to a state concerned about the country.
	education (529)	public (5), country (4), root (3)	Public education needs to be changed because it has many problems.
	health (276)	public (23), country (2), State (2)	Change public health because it is disgusting and degrading to go to the hospital, where you have to be for hours and hours.
	way (234)	make politics (9), educate (5), have (4)	Change the way of making politics.
improve (5603)	education (628)	public (35), civic (9), municipal (4)	Improve municipal education because the teaching method is bad compared to private education
	system (576)	health (135), pension (65), judicial (24)	Improve and change the pension system.
	health (469)	public (73), mental (15), public (5)	Improve public health.
	access (225)	health (51), education (38), housing (13)	I think we need to improve access to education, since many people have to work and study at the same time, and often there is not enough time.
	pension (199)	elderly (17), grandfather (5), retiree (4)	Improve the pensions of retirees who really need them because they are very low and not enough to survive
keep (810)	solidarity (73)	chilean (15), characterize (3), have (3)	It would maintain the solidarity of Chileans.
	system (30)	afp (7), democratic (2), economic (2)	Maintain the democratic system.
	democracy (27)	freedom (1), total reject (1)	I would like to maintain democracy, total rejection to any dictatorship.
	union (22)	chilean (2), chilean people (2), territorial community (1)	Maintain the union of the Chilean people.
	spirit (17)	solidary (4), fight (2), change(1)	The country must maintain the solidary spirit.

Figure 4. Visualization of selected components, *We have to talk about Chile*.

3.2.2. Main Component

The question "What should be the fundamental rights and freedoms of individuals?" clearly prompts responses in the form of nouns. The main component to extract is thus nominal core-arguments (nouns as direct objects of the sentence). The nominal "right" is the most frequent, followed by "freedom" and "education".

3.2.3. Additional Components

Examining the modifiers associated with the nominal "right" reveals that most of the top modifiers—such as "life", "education", "health"—also appeared as main components. This highlights the significance of these topics in participants' responses, and the extraction of modifiers brings them to the forefront of the analysis.

3.2.4. Visualization

The table in Figure 5 shows the final result of the analysis of this database. As in Figure 4, the extracted components are presented hierarchically. In this case, the main components shown are the five most frequent nominal nouns, their modifiers, and a randomly selected sentence containing each of the nominal nouns and their modifiers. As in the analysis previously shown in Section 3.1, this table provides a direct answer to the question "What should be the fundamental rights and freedoms of individuals?"

right (500)	education (99), life (88), health (83)	Right to freedom, to life, to free expression and to education, especially one's own freedom.
freedom (141)	expression (52), opinion/thought (25), cult (10)	Freedom of expression and diversity.
education (54)	quality (17), free (11), health (2)	Free, quality education and health care.
health (51)	mental (7), physical (4), quality (3)	More accessible and better quality health care.
access (31)	health (11), housing (9), education (6)	This must be guaranteed with access to social security, physical and mental health, education, freedom of expression, housing, etc., on an equal basis.

Figure 5. Visualization of selected components, *Citizen Dialogues*.

This visualization allows us to observe not only the absolute frequency of core components, but also how they relate to their selected complements. The hierarchical structure of the table helps illustrate the distribution of core components (in this example, nominal nouns) and their modifiers, making it easier to identify patterns in how these terms appear together in the dataset. In the example, the component "right" occurred 500 times in the data, but the most frequent complement had 99 instances. This relative frequency should be taken into account when interpreting results, both to contextualize the generalizability of conclusions based on these complements, and to support substantive analysis. For instance, it can inform the analysis of heterogeneity and lexical dispersion.

The methodological decisions underlying the previous results are: (i) the plural of "right" appears in the fifth place and was removed; (ii) the nominal count includes all instances, whether or not an associated modifier is present. For the analysis of modifiers, by contrast, empty fields are not considered.

4. Discussion

4.1. Evaluating the Syntagmatic Decomposition

Given the complexity of language, the application of techniques that rely on syntactic relations faces several challenges. A major limitation comes from the syntagmatic decomposition algorithm. Our method relies on patterns of dependency structures recognized in our corpus, but there are cases that escape these patterns. This pattern identification is a crucial step in our methodology, and a detailed exploration of example sentences helps capture its complexities. Our method follows a general approach, but can be extended to capture more patterns, depending on the question or the type of data. Dependency parsers also make mistakes, as do all pre-trained models (see https://stanfordnlp.github.io/stanza/performance.html, accessed on 1 August 2026) on the performance of the Stanza dependency parser). The parser may confuse adjectives with participial verbs and may have trouble dealing with the subjunctive mood in Spanish.

To provide a quantitative estimate of reliability, we evaluated our algorithm by manually performing the syntagmatic decomposition in a randomly selected subset of 100 sentences per database. For the TQH database, we took a random sample of 100 opinions. Since the *Citizen Dialogues* database may contain, for a single opinion, very short phrases, we took a random sample of opinions with three or more words (which is the minimum in the TQH sample). Then, for each opinion in each database, we analyzed the first sentence/independent clause. A sentence was classified as correctly processed

if the decomposition extracted the main components (subject, verb, object, second object, modifiers). The analysis was carried out separately by two of the co-authors, looking at the decomposition generated by our algorithm. Since we did not intend to create full consensus, we opted to report the assessment of the two coders after investigator triangulation. Thus, the differences in classification between coders were discussed. After discussion, the partial consensus between coders was 89% for TQH and 93% for *Citizen Dialogues*. For TQH, the classification accuracy was 84% for coder 1 and 89% for coder 2. For the *Citizen Dialogues*, it was 77% for coder 1 and 74% for coder 2. The algorithm performs better when the data follow a structured collection methodology, as in *We Have to Talk About Chile*, where trained facilitators produced clearer sentence formulation and better segmentation.

An additional concern is whether the errors made by the algorithm penalize certain types of style or components differently. The manual evaluation of the decomposition showed that the component with the highest number of failures is the nominal modifier—that is, the final elements in our extraction chain (since we first extract verbs, then subject/object, and then their modifiers). In particular, this problem occurs in the *Citizen Dialogues* dataset, where sentences contain many conjugated elements and the parsers do not correctly assess whether some elements are objects or modifiers. In the evaluation of this dataset, 11 of 23 classification mistakes are errors of modifier extraction (coder 1). In the TQH evaluation, however, sentences with many conjugated elements are not common, and the problem here is different. Very complex, poorly written sentences cause the parser to fail to parse the sentence overall, in which case no component is extracted correctly. This is the case for 8 of 11 classification mistakes in the TQH evaluation for coder 2.

4.2. Can LLMs Perform Syntagmatic Decomposition?

Working with texts makes it inevitable to consider using LLMs. However, the use of LLMs in this type of application raises concerns about several aspects. The first aspect is the internal explainability of the model, that is, the possibility of examining the mechanism by which a neural network arrives at a given prediction. The literature on interpretability in machine learning has consistently noted that complex models operate as black boxes, and that the explanations available tend to be post-hoc approximations rather than faithful descriptions of the internal process [36,37]. Recent empirical work, specifically on large-scale language models, confirms that this limitation persists even when the model produces seemingly transparent explanations: chain-of-thought explanations can be systematically unfaithful to the actual process that led to the prediction [38], and explainability methods for LLMs under the prompting paradigm face their own methodological challenges, distinct from those of the traditional fine-tuning paradigm [39].

The second aspect concerns provenance and traceability at the system level, that is, the ability of the workflow surrounding the model to record which source sentence each extraction comes from, what prompt was used and under which configuration. There are already mechanisms aimed at this: ALCE [40], the first benchmark designed to measure LLMs' ability to generate text with verifiable, sentence-level citations, or architectures such as PaperTrail [41], which decompose both the generated response and the source documents into claims and verifiable evidence. A third aspect is reproducibility, understood as the stability of the results produced by the system when the same input is run repeatedly under the same configuration. Unlike system-level traceability, which depends on the design of the pipeline that encompasses the model, reproducibility depends on the model's own behavior with respect to sampling parameters such as temperature.

To test these distinctions in practice and evaluate whether an LLM can perform syntagmatic decomposition, we conducted an additional comparison. GPT-5.6 Sol was

accessed through the OpenAI Responses API, using medium reasoning effort, to perform the same structured decomposition task carried out by our method, on the same sample of 100 responses from the TQH database previously used for evaluation. The prompt instructions are provided in the Supplementary Material. The decomposition results were manually validated using the same two-coder strategy reported in the previous section. The accuracy of the prediction was 84% for coder 1 and 80% for coder 2. As can be seen, the LLM performs similarly to our algorithm in the classification task. In order to obtain a satisfactory decomposition, it was necessary to refine the prompt using successful examples. The initial instructions for identifying the subject, verb, and object did not yield an acceptable decomposition. Therefore, we used a set of examples that were correctly classified during the evaluation stage to show the LLM what it was expected to do. These examples were used as demonstrations for the model and were excluded from the subsequent evaluation.

We also examined the stability of the LLM outputs across repeated runs. Because the GPT-5 API does not allow the temperature parameter to be set, we performed three identical runs, keeping the model, prompt and reasoning effort fixed. For each syntactic component, stability was calculated as the proportion of responses for which the extracted value was identical in all three runs. A decomposition was considered fully stable only when all the extracted components were identical across the three runs. The results are summarized in Table S1 of the Supplementary Material. At a component level, the extracted syntactic elements showed generally high stability, the lowest values being 86% for both the second nominal object and its modifier, and achieving 94% and 95% at their highest for the verb and the subject respectively. The fraction of fully stable decompositions was lower (65%), as this was a stricter criterion. Thus, relatively small variations in individual components accumulated at the level of the full decomposition. This pattern is consistent with previous reports showing that even with temperature fixed at 0, the exact agreement rate across repeated runs of an LLM can drop to zero for a substantial share of the tasks evaluated, since temperature does not guarantee determinism in practice [42].

In sum, the LLM achieved accuracy comparable to the proposed parser-based approach on this evaluation sample. As we documented and reported our prompt, system-level traceability is also under control. Internal explainability is always an issue with neural networks, but our dependency-parser algorithm is also based on language models trained with neural networks. In terms of reproducibility, even when the LLM results are not deterministic, extracted syntactic elements showed generally high stability. In any case, whether syntagmatic decomposition is carried out by dependency parsers or via an LLM does not change one of the fundamental premises of the method we propose: that the selection and identification of syntactic components can yield simple, explanatory reconstructions of public opinions. In that sense, the method goes beyond decomposing phrases: it begins by recognizing which components are the most informative given the type of response that a question elicits.

4.3. Comparison Between Datasets and the Role of Facilitators

The accuracy gap between the two datasets is best read not as an isolated empirical finding but as a corroboration that the design of the participation protocol shapes the processability of citizen input, and that this design choice tracks equity considerations as well as computational ones. The explanatory link can be traced directly to the formatting differences already noted in this section. In TQH, facilitators recorded each contribution as a complete grammatical sentence and validated it with the participant before moving to the next question, which standardizes clause structure and punctuation at the point of capture. In the *Citizen Dialogues* dataset, participants self-recorded answers without a shared format,

frequently as itemized lists with inconsistent punctuation, numbering, symbols, and line breaks not necessarily aligned with sentence boundaries. These are precisely the features that our clause-segmentation step struggles with.

Beyond processability, the TQH protocol also incorporated an online validation step in which the facilitator's rendering of each contribution was confirmed with the participant before being recorded. This step can be characterized as a form of member checking [34]. This distinction suggests, as a hypothesis rather than an established finding, that facilitated and participant-validated annotation may compensate for asymmetries in linguistic capital rather than reproduce them, in a manner consistent with the equal opportunity for expression that Habermas's theory of communicative action requires. This reading needs to be qualified, however. Experimental evidence shows that even minimally expressed positions from moderators can shift participants' attitudes and behavior in deliberative settings, and the qualitative methods literature cautions that participant validation can amount to a routine confirmation rather than a substantive check on accuracy [43]. Without an independent assessment of facilitator fidelity, comparable to protocols developed for public deliberation [44], we cannot rule out that part of the accuracy gain in TQH reflects facilitators imposing their own syntactic and lexical choices rather than eliciting the participant's own formulation more faithfully.

The fact that TQH's validation occurred in front of the deliberating group, rather than privately between facilitator and speaker, offers a partial response to this concern. Any participant who had heard the original contribution could contest the facilitator's paraphrase, which reduces, though does not eliminate, the risk that confirmation reflects deference to the facilitator rather than an accurate rendering of the original statement. A limitation of this analysis is that we do not empirically evaluate the reliability of group-witnessed validation against alternative annotation strategies. We consider group validation conceptually preferable to private respondent validation, for the reasons stated above, but this coherence remains a normative argument rather than a demonstrated empirical advantage. A formal evaluation comparing the efficiency and reliability of group-based validation against AI-assisted individual writing support (see, for example, Jungherr and Rauchfleisch, 2025 [45]) remains an open task for future work.

4.4. On the Applicability to Other Languages

A final concern that must be addressed is the applicability of this method to other languages. As we mentioned previously, the syntagmatic decomposition algorithm was designed for the Spanish language. All the figures in the result section were obtained with our Spanish database and translated just for visualization. Likewise, the performance of the method was obtained for the Spanish results (see Section 4). However, the English sentences shown in Table 2 were selected to illustrate that the algorithm's logic operates correctly on a curated set of examples, and should not be read as a performance evaluation.

The claim about ease of adaptation concerns the algorithm's dependence on part-of-speech categories and dependency relations, which are attested across most languages, not on any expectation about comparative accuracy [46,47]. Since most languages differ primarily in word order and in a limited set of typological parameters, extending the algorithm to a new language should mainly involve adjusting how modifiers and core arguments are located relative to the head, rather than redesigning the decomposition logic itself. Whether this structural adaptability translates into comparable or superior performance in languages with less flexible word order than Spanish remains an open empirical question, which we leave for future work with a proper evaluation dataset.

4.5. Comparison Against Other Methods

The database from *We have to talk about Chile* contains 38,606 responses for the question "What should we change, improve, or maintain in Chile?" Thus, a suitable approach for comparison could be topic modeling, which can deal with large databases. Here, we implemented BERTopic, a topic modeling technique that combines a sentence transformer model with a clustering algorithm to generate topic representations with a class-based TF-IDF procedure [48]. Since the purpose of BERTopic differs from that of the proposed method, we present this comparison primarily as a qualitative rather than a quantitative benchmark.

Our implementation used a sentence transformer model in the Spanish language (hiiamsid/sentence _similarity _spanish _es) and default algorithms to reduce dimensionality (UMAP, UMAP_N = 5, UMAP_DIST = 0.01) and clustering (HDBSCAN, MIN_CLUSTER_SIZE = 100). The minimum cluster size was set to 100 to avoid clusters that were too small while keeping a relatively reduced number of clusters. Topic representations were generated by filtering out overly common words from the vocabulary (max_df = 0.8) and stopwords, and weighting the terms using a class-based TF-IDF scheme (bm25_weighting). The number of outlier documents initially reached 23%, so we also applied an outlier reduction step. Under the aforementioned parameters, the model identified 30 topics.

The full results of the BERTopic model can be found in the Supplementary Material, including, for each topic, the number of documents, 10 representative words, and one representative document. In general terms, we can say that with 30 topics, the model manages to identify fairly specific themes. Thus, for example, in addition to the corpus's major themes, such as education (topic 1) or health (topic 2), the model also recognizes topics related to immigration (topic 28), probity (topic 24), and citizen participation (topic 22).

As can happen with any algorithm that performs topic modeling, there may be clusters that are difficult to interpret or distinguish from other clusters. This is the case, for example, of topic 15, which seems to be confused with the education topic (topic 1), because its representative terms include words such as "educational", "to study", "schools". But, in general terms, the model successfully identifies the latent topics. Therefore, the value of this type of model lies in identifying the topics discussed in the corpus and accounting for the relevance of each topic (based on the number of documents assigned to each topic). Faced with a large corpus, this is a first step that makes it possible to organize the information.

However, beyond identifying topics, the only way to get closer to what is being said is by looking at the representative sentences. The peculiarities of each topic may remain hidden behind the most representative words of the topic. In particular, for the TQH corpus, the model does not tell us what should be improved, changed, or maintained, because the model extracts topics, not the relationships between those topics and those three verbs. One exception is topic 30, which, among its most representative words, includes a verb (solidarity, Chilean, maintain, unity, resilience).

The *Citizen Dialogues* database, on the other hand, is already separated by topic and contains fewer observations. Thus, a bigram semantic network is a suitable alternative for comparison. This type of visualization has already been used in several applications [27–29], even with Chilean data [49]. Bigram extraction is more informative than single-word analysis, and networks make it possible to understand the associations between them. To build the network, bigrams were extracted from all the texts. The preprocessing consisted solely of converting the text to lowercase. Only bigrams that match the noun–adjective pattern were used, ensuring that each bigram contains a nominal plus a modifier. Then, we built an undirected and weighted network, where each word is a node, and the weight of the link between words reflects the number of times that bigram appears in the corpus. To keep the network size limited and avoid unconnected dyads, we used

only bigrams with 3 or more occurrences. The figure showing the network is included in the Supplementary Material.

The bigram semantic network and the predicative component analysis both allow readers to address what was discussed during the *Citizen Dialogues*. However, they afford and obstruct different kinds of interpretative analysis. The visualization of predicative components (see Figure 5) provides a clearer account of the dominant concerns of the dialogues, identifying rights, freedom, education, health and access as prominent themes and showing the qualities or demands associated with each through the modifiers. It foregrounds the relative importance of these themes and translates dispersed contributions into a concise structure that allows the inference of main claims, such as demands for free, high-quality education or more accessible healthcare. This is further associated with representative examples that also facilitate traceability to the original text.

Yet, this clarity comes at the cost of reducing the text's diversity and complexity. By selecting only a handful of components and synthesizing them into a simplified structure, our visualization can make the contributions of the participants appear more unified and settled than they may have been. It is also less apt to visualize the less frequent aspects of the corpus. The bigram semantic network, by contrast, makes the relational structure of the dialogue more visible. Rather than isolating a small number of dominant themes, the network shows how concepts such as rights, life, health, education, dignity, freedom, social protection and public provision connect to each other. The network also retains less frequent associations, including drinking water, environmental conditions, decent work, migration status and the needs of older people, which are left out by our proposed method.

However, at the same time, its visual density makes the network considerably harder to read, while the connections alone cannot establish the precise meaning, direction, or context of an association. In that sense, it allows readers to grasp the words and concepts that are central to the corpus, but says very little about them beyond which ones are more central/peripheral or notable connections such as "fundamental" and "rights" or "free" and "education", which one can easily understand as key bigrams. Here, traceability is also harder to construct, given the larger number of elements and frequencies that are represented.

The comparison therefore demonstrates that each method actively shapes the analysis's affordances. Our visualization of predicative components affords thematic prioritization, readability and communicable claims, but obstructs attention to ambiguity, peripheral concerns, and the interdependence of concepts. The semantic network provides exploratory and relational interpretation, but obstructs deeper or more straightforward conclusions about what participants collectively prioritized or intended.

4.6. The Problem of Epistemic Exclusion

One of the challenges faced by any systematization method, in light of the normative criteria previously set out, is epistemic exclusion. Models that prioritize more complex, information-rich opinions may amplify the voices of those who already have greater representation in the public sphere, reproducing the cultural capital inequalities that deliberative design attempts to correct.

One way to test whether our method preserves epistemic inclusion is to see whether the algorithm performs better on longer sentences, which presumably contain more information. When we carried out the component-by-component evaluation (see Section 4.1), we realized that in very complex and poorly written sentences, i.e., sentences containing grammatical errors, the parser could not correctly evaluate the sentence structure and the syntagmatic decomposition failed. Thus, when the language model does not understand the grammatical structure, the problems of epistemic exclusion are reproduced.

However, long but well-structured sentences do not have that problem. If the grammatical structure is correctly captured, the length of the sentence does not matter, so long as it contains the components one wants to capture. In this sense, (well-written) sentences that contain more information will not carry any additional weight, since what matters are the grammatical components they contain. This reinforces the importance of the facilitator in taking notes during the dialogue.

However, this analysis does not constitute sufficient empirical evidence to claim whether or not our method incurs epistemic exclusion. Future analyses should use better metrics of linguistic complexity and a larger number of samples to statistically evaluate the precision of the model in sentences of different levels of complexity.

5. Conclusions

This article began with a practical problem—how to automatically process the opinions produced in large-scale deliberative citizen participation processes—and arrived at an argument that goes beyond the technical. The automated processing of citizen participation is a problem of normative design whose solution must be coherent with the philosophical, ethical, and institutional decisions that the participatory process embodies. In this regard, we identified two challenges in automated citizen participation processing, which go beyond the inefficacy of traditional methods in handling large quantities of textual data. The first is the risk of epistemic exclusion, which reproduces the cultural capital inequalities that deliberative design attempts to correct, and threatens the emancipatory principle of participation [5]. Second, national and international regulatory frameworks establish that AI systems used in democratic governance contexts must be transparent, explainable, auditable, and coherent with the aims of the process they systematize.

This article explored a novel approach employing syntactic analysis to navigate these socio-technical challenges, particularly in the context of large-scale participation where the conventional tools of textual analysis and Natural Language Processing (NLP) fall short. Our method shifts focus from word counts—like Wordclouds or Topic Modeling—to a lexical–syntactic analysis, which dissects sentences into their syntactic components. This approach allows for a more nuanced reconstruction of citizen ideas within the context of whole ideas structured in actual sentences, rather than the impressionistic picture that is produced through words and concepts alone. This, of course, rests on the assumption that the grammatical structure of the ideas people utter is an adequate representation of their thinking.

By leveraging advances in PoS taggers and dependency parsers, we propose a four-stage analysis method in which the analyst selects the core components and terms to anchor the visualization and then selects the additional components that best reconstruct the ideas. Finally, this sentence reconstruction employs a structured visual representation where font size indicates frequency (as in the case of Wordclouds), but the context is maintained through associated modifiers and examples, thus offering both breadth and depth in understanding public discourse. The lexical–syntactic approach can also be adjusted to the needs of the problem, depending on whether the question elicits a greater use of verbs, objects, or adjectives. Unlike methods that require training, our approach can work with a reduced amount of data, since it makes use of pre-trained language models.

Returning to the framework of Bächtiger and Parkinson (2019), our method contributes to the normative goals of participation in different ways [5]. With respect to epistemic goals, lexical–syntactic analysis identifies not only the most frequent topics but the predicative structure of citizen positions on those topics, offering a reconstruction of collective intelligence that word-frequency methods cannot provide. With respect to legitimacy goals, the traceability of the method satisfies the publicity and accountability requirements that

democratic legitimacy demands of the systematization process. Regarding emancipatory goals, our method does not exclude contributions prior to the analysis based on estimated informativeness, so the voices of participants with less linguistic capital are still included in the analysis. However, since our method still involves frequency-based presentation, term selection, and decisions about which components are visualized, full corpus coverage should not be equated with equal representation in the final analytical output.

We have highlighted the limitations of our method. The algorithm generally performs well on standard text but struggles with complex linguistic structures. All these considerations make the data collection process particularly relevant in a citizen dialogue design or, broadly, in any instance where opinions are generated in text form. As we saw in the results section, the algorithm's performance was higher when the data came from TQH, where trained facilitators took notes following a grammatically standard format. Facilitators also fulfill other functions recommended by deliberative theory, such as ensuring respect among participants and regulating interventions. Future participatory designs should take all these aspects into account to help safeguard the quality of the process.

Overall, this study presents a compelling alternative for interpreting large volumes of citizen input. Using syntactic analysis, it offers a pathway toward a more meaningful reconstruction of citizen opinions beyond word frequencies. Our contribution is not the syntactic techniques themselves, which are well-known, nor the syntagmatic decomposition algorithm itself. As we saw, an LLM was also able to carry out that task. What we propose is a model that integrates these techniques into a framework explicitly designed around the normative requirements of citizen participation. Future research could explore further refinement of these tools, expanding their application to more languages and participatory contexts, thereby fostering a more robust dialogue between citizens and governance structures.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/computation14090209/s1, Figure S1: LLM prompt for syntagmatic decomposition; Table S1: Output stability across three identical LLM runs for the TQH dataset; Table S2: BERTopic results; Figure S2: Bigram Network for the Citizen Dialogues database.

Author Contributions: Conceptualization, M.P.R. and J.I.G.; methodology, M.P.R. and P.A.; software, P.A.; writing—original draft preparation, M.P.R., J.I.G. and P.A.; writing—review and editing, C.F.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting the findings of this study are available upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Berliner, D. Information Processing in Participatory Governance. *SocArXiv* **2023**. [CrossRef]
2. Chambers, S.; Warren, M.E. Why deliberation and voting belong together. *Res. Publica* **2023**, *31*, 279–297. [CrossRef]
3. Lafont, C. Can Democracy be Deliberative & Participatory? The Democratic Case for Political Uses of Mini-Publics. *Daedalus* **2017**, *146*, 85–105. [CrossRef]
4. Niemeyer, S. Deliberation and ecological democracy: From citizen to global system. *J. Environ. Policy Plan.* **2020**, *22*, 16–29. [CrossRef]
5. Bächtiger, A.; Parkinson, J. Unpacking Deliberation. In *Mapping and Measuring Deliberation*; Oxford University Press: Oxford, UK, 2019; pp. 19–44. [CrossRef]

6. Goñi, J.I.; Fuentes, C.; Raveau, M.P. An experiential account of a large-scale interdisciplinary data analysis of public engagement. *AI Soc.* **2023**, *38*, 581–593. [CrossRef]
7. Porter, T.M.; Nelson Espeland, W. Data and Expertise: Some Unanticipated Outcomes. In *The Oxford Handbook of Expertise and Democratic Politics*; Eyal, G., Medvetz, T., Eds.; Oxford University Press: Oxford, UK, 2023; pp. 258–281. [CrossRef]
8. Russell, S.; Williams, R. Social Shaping of Technology: Frameworks, Findings and Implications for Policy, with Glossary of Social Shaping Concepts. In *Shaping Technology, Guiding Policy*; Sørensen, K., Williams, R., Eds.; Edward Elgar: Cheltenham, UK, 2002; pp. 37–132.
9. Romberg, J.; Escher, T. Making Sense of Citizens' Input through Artificial Intelligence: A Review of Methods for Computational Text Analysis to Support the Evaluation of Contributions in Public Participation. *Digit. Gov. Res. Pract.* **2024**, *5*, 1–30. [CrossRef]
10. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
11. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1532–1543.
12. Devlin, J. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
13. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762.
14. Confalonieri, R.; Coba, L.; Wagner, B.; Besold, T.R. A historical perspective of explainable Artificial Intelligence. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2021**, *11*, e1391. [CrossRef]
15. UNESCO. *Recommendation on the Ethics of Artificial Intelligence*; UNESCO: Paris, France, 2021.
16. OECD. *Recommendation of the Council on Artificial Intelligence*; OECD/LEGAL/0449; OECD: Paris, France, 2019.
17. MCTIC. *Política Nacional de Inteligencia Artificial*; MCTIC: Santiago, Chile, 2025.
18. BCN. *Proyecto de Ley sobre Inteligencia Artificial*; Boletín N°16821-19; BCN: Santiago, Chile, 2024.
19. Bourdieu, P. *Language and Symbolic Power*; Harvard University Press: Cambridge, MA, USA, 1991.
20. Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 610–623.
21. Young, I.M. *Inclusion and Democracy*; OUP Oxford: Oxford, UK, 2000.
22. Fraser, N. *Scales of Justice: Reimagining Political Space in a Globalizing World*; Columbia University Press: New York, NY, USA, 2009; Volume 31.
23. Tenemos que Hablar de Chile. *Un País que se Piensa y Proyecta: Diez Hallazgos desde un Chile a Escala*, 1st ed.; Tenemos que Hablar de Chile: Santiago, Chile, 2021.
24. Naseem, U.; Razzak, I.; Khan, S.K.; Prasad, M. A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models. *ACM Trans. Asian-Low-Resour. Lang. Inf. Process.* **2021**, *20*, 1–35. [CrossRef]
25. Vilenchik, D. Simple statistics are sometime too simple: A case study in social media data. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 402–408. [CrossRef]
26. Dang, E.K.F.; Luk, R.W.P.; Allan, J. Beyond bag-of-words: Bigram-enhanced context-dependent term weights. *J. Assoc. Inf. Sci. Technol.* **2014**, *65*, 1134–1148. [CrossRef]
27. Green, E.P.; Whitcomb, A.; Kahumbura, C.; Rosen, J.G.; Goyal, S.; Achieng, D.; Bellows, B. "What is the best method of family planning for me?": A text mining analysis of messages between users and agents of a digital health service in Kenya. *Gates Open Res.* **2019**, *3*, 1475. [CrossRef] [PubMed]
28. North, S.; Piwek, L.; Joinson, A. Battle for Britain: Analyzing events as drivers of political tribalism in Twitter discussions of Brexit. *Policy Internet* **2021**, *13*, 185–208. [CrossRef]
29. Valença, G.; Moura, F.; de Sá, A.M. How can we develop road space allocation solutions for smart cities using emerging information technologies? A review using text mining. *Int. J. Inf. Manag. Data Insights* **2023**, *3*, 100150. [CrossRef]
30. Wattenberg, M.; Viégas, F.B. The word tree, an interactive visual concordance. *IEEE Trans. Vis. Comput. Graph.* **2008**, *14*, 1221–1228. [CrossRef] [PubMed]
31. Jurafsky, D.; Martin, J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd ed.; Online Manuscript Released January 12, 2025. 2025. Available online: https://web.stanford.edu/~jurafsky/slp3/ (accessed on 1 August 2026).
32. De Marneffe, M.C.; Manning, C.D.; Nivre, J.; Zeman, D. Universal dependencies. *Comput. Linguist.* **2021**, *47*, 255–308. [CrossRef]
33. Qi, P.; Zhang, Y.; Zhang, Y.; Bolton, J.; Manning, C.D. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 101–108.
34. Lincoln, Y.S.; Guba, E.G. *Naturalistic Inquiry*; Sage: Newbury Park, CA, USA, 1985.

35. SEGPRES. *Participación Ciudadana en el Proceso Constitucional 2023: Diálogos Ciudadanos Autoconvocados*; Technical report; Secretaría de Participación Ciudadana: Santiago, Chile, 2023.
36. Lipton, Z.C. The mythos of model interpretability. *Commun. ACM* **2018**, *61*, 36–43. [CrossRef]
37. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef] [PubMed]
38. Turpin, M.; Michael, J.; Perez, E.; Bowman, S. Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 74952–74965. [CrossRef]
39. Zhao, H.; Chen, H.; Yang, F.; Liu, N.; Deng, H.; Cai, H.; Wang, S.; Yin, D.; Du, M. Explainability for large language models: A survey. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 1–38. [CrossRef]
40. Gao, T.; Yen, H.; Yu, J.; Chen, D. Enabling large language models to generate text with citations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 6465–6488.
41. Martin-Boyle, A.; Leckey, C.; Brown, M.; Kaur, H. Papertrail: A claim-evidence interface for grounding provenance in llm-based scholarly q&a. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2026; pp. 1–25.
42. Ouyang, S.; Zhang, J.M.; Harman, M.; Wang, M. An empirical study of the non-determinism of chatgpt in code generation. *ACM Trans. Softw. Eng. Methodol.* **2025**, *34*, 1–28. [CrossRef]
43. Birt, L.; Scott, S.; Cavers, D.; Campbell, C.; Walter, F. Member checking: A tool to enhance trustworthiness or merely a nod to validation? *Qual. Health Res.* **2016**, *26*, 1802–1811. [PubMed]
44. Draucker, C.; Carrión, A.; Ott, M.A.; Knopf, A. Assessing facilitator fidelity to principles of public deliberation: Tutorial. *JMIR Form. Res.* **2023**, *7*, e51202. [CrossRef] [PubMed]
45. Jungherr, A.; Rauchfleisch, A. Artificial Intelligence in deliberation: The AI penalty and the emergence of a new deliberative divide. *Gov. Inf. Q.* **2025**, *42*, 102079. [CrossRef]
46. Dryer, M.S. Word order. In *Language Typology and Syntactic Description*; Shopen, T., Ed.; Cambridge University Press: Cambridge, UK, 2007; pp. 61–131.
47. Schachter, P.; Shopen, T. Parts-of-speech systems. In *Language Typology and Syntactic Description*; Shopen, T., Ed.; Cambridge University Press: Cambridge, UK, 2007; pp. 1–60.
48. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
49. Mascareño, A.; Cordero, R.; Henríquez, P.; Ruz, G. Comunidades semánticas y distinciones políticas en el discurso de los convencionales constituyentes. *Puntos Ref.* **2021**, *581*, 1–27.