CAMBRIDGE
UNIVERSITY PRESS

ARTICLE TYPE

Expanding Your Vocabulary: A Framework for Topic Integration in Texts

Roy Gardner,[†‡] Matthew Martin,[*¶‡] Ashley Moran,[¶‡] Zachary Elkins,[¶‡] Andrés Cruz,[¶‡] and Guillermo Pérez[¶‡]

†PeaceRep, University of Edinburgh, Edinburgh, EH8 9YL, UK
‡Comparative Constitutions Project, Austin, 78712, Texas, USA
¶Department of Government, University of Texas at Austin, Austin, 78712, Texas, USA
*Corresponding author. Email: mjmartin@utexas.edu

Abstract

Topic discovery and integration are vital for maintaining vocabularies that categorize textual corpora. Automated approaches are often computationally expensive and lack domain-specific conceptual nuance; manual approaches are costly in terms of time and potential bias. To address this dilemma, we introduce the segments-as-topic (SAT) methodology, a four-stage process that combines automation and human expertise to assess candidate topics for vocabulary inclusion. In the SAT generation stage, a topic is formulated and refined through collaboration with domain experts, then a sentence-level semantic similarity model retrieves corpus segments semantically aligned with the topic. The SAT expansion stage uses this seed set to find additional semantically similar segments, which are iteratively accepted or rejected to build a final segment set. During the review stage, a panel of scholars evaluates the topic for inclusion. In the integration stage, all segments in the final segment set are automatically tagged with the new topic. We apply this methodology to the Comparative Constitutions Project vocabulary that tracks over 330 topics in national constitutions, and demonstrate the addition of three new topics to the vocabulary. The SAT approach balances computational efficiency with expert judgment, offering a systematic, user-friendly, and replicable framework for social scientists to expand domain-specific vocabularies.

Keywords: topic integration; vocabularies; semantic similarity; constitutions

1. Introduction

“The limits of my language mean the limits of my world,” the Austrian philosopher Ludwig Wittgenstein (1922) observed more than a century ago. This statement captures the relationship between language and perception—our language shapes our understanding of the world around us.[1] For social scientists, this relationship is critical when connecting evolving corpora with conceptual vocabularies (closely related to a taxonomy, schema, and ontology) that represent domain knowledge. Robust vocabularies must evolve to accommodate emerging topics from new corpus additions, paradigm shifts, or cross-disciplinary influences. If we then interpolate Wittgenstein’s dictum, the limits of our scholarly vocabularies mean the limits of our domain of study.

Political scientists have wrestled with vocabulary curation at least since Giovanni Sartori’s (1984) influential work on concepts. Sartori recognized that concepts are critical for representing knowledge and that coordination among scholars about vocabulary advances scientific discovery. Building on this foundation, subsequent contributions—such as Collier and Mahon’s (1993) distinction between

[1] The influence of language on cognition has been debated among linguists since Whorf’s (2012 [1956]) work on linguistic relativity.

classical and family resemblance approaches, and Gerring's (2011) and Goertz's (2020) works on concept structure and measurement—have further refined how political scientists think about conceptual validity and the relationship between concept and indicators.

The advance of computational methods provides significant leverage for the organization and use of concepts. One core challenge is to update conceptual vocabularies and apply them to related examples in text. This paper presents a methodology to formulate, evaluate, and incorporate new topics into existing vocabularies. The assumed context is a domain in which a core set of texts represents domain knowledge. We focus on constitutional law and a corpus of world constitutions as our exemplar domain. Our framework employs a semantic similarity model, which represents sentences as vectors of numbers, to encode existing topics, candidate topics, and relevant text segments (Cruz et al. 2023; Gardner 2023). We develop the segments-as-topic (SAT) approach that uses corpus segments to represent candidate topics, then identifies additional similar segments to define the topic's conceptual scope. Throughout this process, domain experts actively participate in topic refinement rather than being limited to post hoc validation.

Our methodology comprises four stages: (1) SAT generation; (2) SAT expansion; (3) SAT review; and (4) SAT integration into the corpus. We demonstrate this process using vocabulary that authors of the Comparative Constitutions Project (CCP) developed to track topics within a corpus of national constitutions (Elkins and Ginsburg, 2025 [2005]). The CCP's indexed repository of constitutional texts, Constitute, includes 330 core topics and is widely used by scholars and constitutional drafters. As such, the integration of new topics is potentially consequential for future constitutional development.

This approach demonstrates the synergy between natural language processing and human expertise, producing topics that are both semantically coherent and resonant with domain experts. We seek to empower curators—scholars and practitioners alike—to identify and develop candidate topics that enrich their own vocabularies. In other words, our tools, which will be publicly available, are designed for application beyond the constitutional domain.

2. The Problem

What we describe is a general concern for scholars studying ideas within specific corpora. Our running example involves national constitutions, but similar challenges exist in other scholarly projects. The Policy Agendas Project (Jones et al. 2023) inventories policy ideas in national legislation, and the Party Manifesto Project (Lehmann et al. 2022) codes election platforms of major political parties across countries with different vocabularies. Beyond these political science examples, comparable challenges exist in analyzing news, social media, books, film scripts, music lyrics, or academic articles. For any such corpus, scholars often aim to develop explicit topic sets to track ideas in the genre—what some call "conceptual ecology" (Cruz et al. 2023, 19). Our focus is how to systematically update such vocabularies in evolving genres.

Maintaining and updating vocabularies presents significant challenges. When corpus expansion reveals the need for vocabulary refinement, researchers often hesitate due to the daunting nature of the task. Evaluating potential vocabulary modifications requires examining the entire corpus to identify segments that would require recoding under a new schema—a prohibitively time-consuming process for substantial corpora.

Researchers often resort to shortcuts rather than comprehensive recoding. One common approach involves searching the corpus using keywords or embeddings to identify potential matches. However, this method operates in isolation from existing topic coding and fails to contextualize proposed modifications within the established vocabulary framework. More critically, the quality of results depends entirely on the search query, which inevitably produces both false positives and false negatives while creating a "moving target" problem when parameters are modified.

An alternative approach leverages topic modeling (TM) to discover latent topics absent from the current vocabulary. Traditional models like LDA (Blei, Ng, and Jordan 2003) or STM (Roberts et al. 2013) typically rely on “bag of words" frameworks, while recent advances incorporate embeddings to better capture contextual nuances (Dieng, Ruiz, and Blei 2020; Grootendorst 2022). KeyATM (Eshima, Imai, and Sasaki 2024) presents a promising deductive approach—a guided algorithm that retrieves specified topics through keywords while allowing discovery of new topics. Despite their appeal, automated TM approaches remain insufficiently robust for effective topic curation, requiring extensive human validation even for simple classification tasks (Ying, Montgomery, and Stewart 2022).

Given these limitations, topic curation in practice typically follows an eclectic but informal procedure combining manual revision, automated search and recoding, and expert judgment. We contend that the difficulty of this process inhibits experimentation with new or revised topics, resulting in undue inertia in vocabulary development. We propose a method that streamlines this process, facilitating and accelerating vocabulary enrichment.

3. Methodological Framework

We develop an approach to topic curation that uses semantic similarity tools to gather a set of segments in our corpus that represents the topic we want to add to our vocabulary. We then leverage these segments to find additional similar segments in the corpus, ultimately identifying the full set of segments to which the topic applies and allowing us to seamlessly integrate the new topic into our vocabulary and corpus. A summary of the framework is in Figure 1. We detail the steps in the following sections.

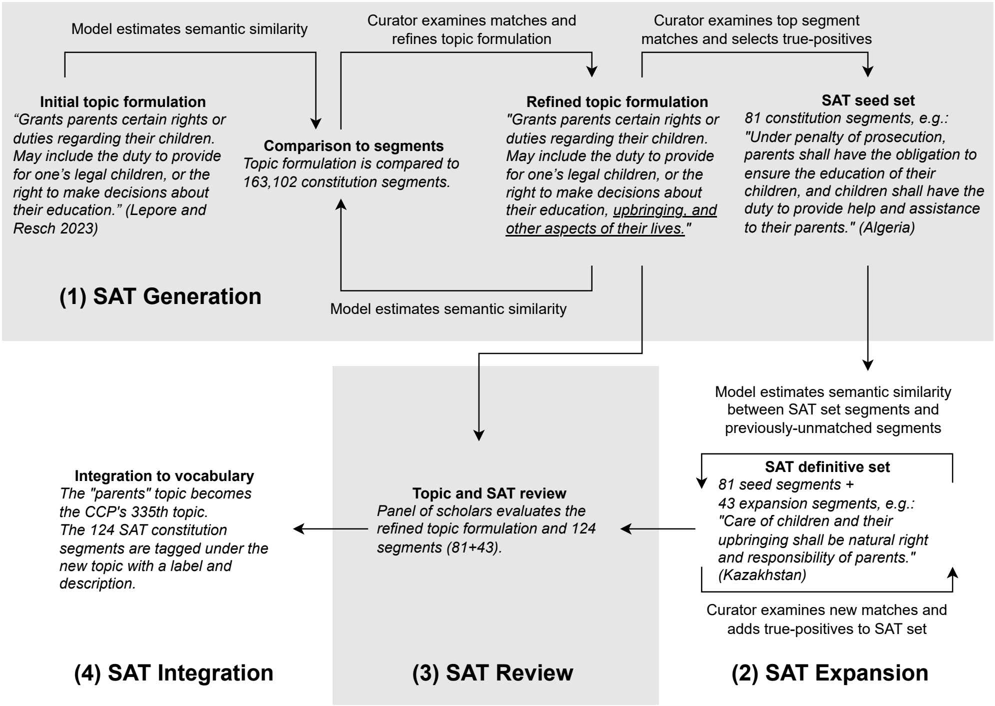

Model estimates semantic similarity
Curator examines matches and refines topic formulation
Curator examines top segment matches and selects true-positives
Initial topic formulation
“Grants parents certain rights or duties regarding their children. May include the duty to provide for one’s legal children, or the right to make decisions about their education.” (Lepore and Resch 2023)
Comparison to segments
Topic formulation is compared to 163,102 constitution segments.
Refined topic formulation
"Grants parents certain rights or duties regarding their children. May include the duty to provide for one’s legal children, or the right to make decisions about their education, upbringing, and other aspects of their lives."
SAT seed set
81 constitution segments, e.g.: "Under penalty of prosecution, parents shall have the obligation to ensure the education of their children, and children shall have the duty to provide help and assistance to their parents." (Algeria)
(1) SAT Generation
Model estimates semantic similarity
Model estimates semantic similarity between SAT set segments and previously-unmatched segments
SAT definitive set
81 seed segments + 43 expansion segments, e.g.: "Care of children and their upbringing shall be natural right and responsibility of parents." (Kazakhstan)
Curator examines new matches and adds true-positives to SAT set
Integration to vocabulary
The "parents" topic becomes the CCP's 335th topic. The 124 SAT constitution segments are tagged under the new topic with a label and description.
Topic and SAT review
Panel of scholars evaluates the refined topic formulation and 124 segments (81+43).
(4) SAT Integration
(3) SAT Review
(2) SAT Expansion

Figure 1. Diagram of the SAT approach to topic integration. Examples from the application are in italics.

3.1 Measuring Semantic Similarity

Sentence-level semantic similarity measures the degree to which two or more natural language sentences or clauses convey similar meaning. This approach has been applied to a range of tasks including text search (Farouk 2018) and machine translation (Yang et al. 2019). Among the methods used to calculate sentence-level similarity, those that represent both the meaning and order of words—known as sentence-sequence representations—show significant promise (Cruz et al. 2023; Gardner 2023; Bestvater and Monroe 2023). While "bag of words" and word-level representations can be aggregated to the sentence level (Rodriguez and Spirling 2022), sentence-sequence representations natively account for both the meaning of individual words and their sequential relationships within a sentence (Aggarwal 2022). This approach captures the context of the natural language in which words appear, allowing the model to recognize subtle differences in meaning that arise from word order or phrasing. These representations are particularly effective in comparing sentences that convey similar ideas but use different vocabulary or structure.

Here we employ version 4 of Google's Universal Sentence Encoder (USE v4) to generate high-dimensional numerical representations of sentences, referred to as encoding vectors or embeddings (Cer et al. 2018). Sentences are represented as discrete points in a 512-dimension semantic space, and the distance between two points is used to measure the divergence in the meaning of the corresponding texts.

The semantic similarity score σ of two text segments *a* and *b* is measured as the inverse of the angular distance between the encoding vectors of the segments. This distance measure performs better on average than cosine similarity (Cer et al. 2018).

$$\sigma(a, b) = 1 - \frac{\arccos\left(\dfrac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|}\right)}{\pi}, \tag{1}$$

where ***a*** and ***b*** are the encoding vectors of *a* and *b* respectively.

The inverse of this distance produces a semantic similarity score ranging from 0.0 to 1.0. A score of 1.0 indicates that two sentences are identical in meaning and comprise the same words in the same order. As the meanings of the sentences diverge, the similarity score decreases; a score of 0.0 indicates completely distinct meanings.

USE models facilitate efficient and accurate computation of sentence-level encoding vectors, enabling large-scale semantic similarity tasks across multilingual datasets with minimal text preprocessing (Cruz et al. 2023; Gardner 2023). Our selection of the version 4 USE model was based on its optimal balance between robust performance on standard benchmarks and computational efficiency, making it well-suited for our analytical requirements. In our own tests, we found that USE version 4 was 70 times faster than USE version 5 when generating encoding vectors, and 30 times faster than SBERT models. However, our methodology is model agnostic in that it is based on the ability to perform sentence-level semantic similarity and is therefore independent of the model generating the encoding vectors. We continuously evaluate emerging models against our speed and accuracy thresholds, and will integrate superior alternatives into our framework as part of our commitment to methodological advancement.

3.2 Data Sources

Our document corpus comprises the text of 192 constitutions in force as of January 2025. CCP has organized these constitutional texts according to their hierarchical structure (sections, subsections, etc.). We flatten the hierarchy, ignore titles and headers, and analyze only those segments containing the substantive content of constitution sections. Altogether, 192 national constitutions provide a total of 163,102 text segments.

3.3 Text Processing

We then process constitution segments for inclusion in our semantic similarity model. This produces a set of indexed identifiers for constitution segments, $S = \{s_1, s_2, \ldots, s_N\}$, as well as corresponding vectors obtained from the USE v4 model for each segment's text, $\boldsymbol{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\}$.

Constitution segment text is stored in a dictionary where each key is a segment identifier. The segment identifier also identifies the segment's constitution and provides access to constitutional metadata.

These constitution segments form the core text against which we test potential new topics for inclusion in the CCP vocabulary. Following these initial preprocessing steps, we begin our segments-as-topics (SAT) approach, using constitution segments to represent and identify a potential new (candidate) topic in our constitutional corpus. Below we detail the four stages of this methodology: SAT generation, SAT expansion, SAT review, and SAT integration into the corpus.

3.4 SAT Generation

SAT generation is an iterative process in which topic formulations—short phrases that capture the meaning of a candidate topic—are tested against the corpus of constitution segments. The output of this process is a small set of constitution segments (the SAT seed set) that best capture the meaning of the candidate topic.

SAT generation involves two steps: (1) measuring the semantic similarity of a topic formulation to text segments in the corpus and identifying segments that are at or above a similarity threshold; and (2) selecting a small set of segments that best match the meaning of the topic–the SAT seed set.

3.4.1 Measuring semantic similarity to constitution segments

Having passed the preliminary test above, the candidate topic's formulation is used to find semantically similar text segments in the constitutions comprising our corpus. This step involves computing similarity scores between the candidate topic and every text segment in the corpus to identify relevant matches.

A vector $\boldsymbol{v}$ is computed that contains the semantic similarity scores between the candidate topic c, and each of the constitution segments in S: $\boldsymbol{v} = \{\sigma(c, s_1), \sigma(c, s_2), \ldots, \sigma(c, s_N)\}$

The curator selects a threshold θ_{search} which is applied to the similarity scores in $\boldsymbol{v}$ to obtain the search results R. R then includes the set of constitution segments at or above that threshold and their semantic similarity scores:

$$(s_n, \boldsymbol{v}_n) \begin{cases} \in R & \textit{if } \boldsymbol{v}_n \geq \theta_{search} \\ \notin R & \textit{if } \boldsymbol{v}_n < \theta_{search} \end{cases} \tag{2}$$

The search threshold determines the degree of similarity and thus also the number of returned search results. The curator can experiment with different search thresholds until they find a threshold that returns a manageable number of search results, i.e., enough results to create a SAT seed set, but not so many that the task of finding the seed set becomes overwhelming. From our experience with constitutions, we found that the optimal search threshold was between 0.62 and 0.68. Employing a relatively high threshold can keep the number of matched segments manageable.

3.4.2 Clustering search results

The search results in R are clustered to facilitate the curator's selection of the seed set. A similarity matrix is constructed where the candidate topic segments in rows and constitution segments in columns map onto the set of segment identifiers in $R = \{r_1, r_2, \ldots, r_K\}$ and their encoding vectors $\boldsymbol{R} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_K\}$.

The similarity matrix $\boldsymbol{W}$ is generated by computing the semantic similarity of every pair of search result segments. The curator applies a threshold $\theta_{cluster}$ to convert the similarity matrix $\boldsymbol{W}$ into a binary-valued matrix $\boldsymbol{H}$ as follows:

$$h_{m,n} = \begin{cases} 0 & \mathit{if}\ w_{m,n} < \theta_{cluster} \\ 1 & \mathit{if}\ w_{m,n} \geq \theta_{cluster} \end{cases} \tag{3}$$

The matrix $\boldsymbol{H}$ represents an undirected graph where the value 1 indicates an at- or above - threshold connection between a pair of search result segments. The connected components of the graph are found using the Python SciPy API. Each connected component identifies a cluster of semantically similar search results. Search results that do not belong to a connected component are referred to as singletons. The search results are organized into cluster and singleton sets before presentation to the curator.

The curator can experiment with the cluster threshold. A low threshold will cluster the results into a few large clusters; a high threshold will cluster the results into many smaller clusters. The curator must select a threshold that best facilitates the process of selecting segments for the seed set. From our work with constitutions we found that the optimal cluster threshold lay between 0.7 and 0.78.

Clustering groups of semantically similar text segments makes it easier for the curator to identify patterns and determine whether a particular topic formulation finds segments that best capture the essence of the candidate topic. The curator can reformulate the topic text and repeat the process. Different formulations can be tested until the curator can select appropriate segments for the SAT seed set.

3.5 SAT Expansion

SAT expansion is also an iterative process. As the name implies, the process expands the SAT set of a candidate topic created in the SAT generation stage. The SAT expansion process, described below, uses the seed set to search the corpus for additional segments that match the meaning of the candidate topic. This represents a change from the SAT generation stage, which used the topic formulation to search the corpus.

3.5.1 Finding new constitution segments

A semantic similarity matrix $\boldsymbol{U}$ is created with the seed set segments in rows and the constitution segments in columns. The next step is to evaluate the similarity scores between the topic seed segments and the constitution segments. This involves analyzing the matrix to identify which constitution segments are semantically similar to the seed segments. Importantly, this discounts the segments already in the seed set and any segments already identified as dissimilar in forming that seed set (now composing rejected set X). A curator-defined threshold θ_{topic} is applied to the similarity scores in $\boldsymbol{U}$ to obtain the SAT search results R – the set of segments at or above that threshold:

$$s_n \begin{cases} \in R & \mathit{if}\ u_{m,n} \geq \theta_{topic} \\ \notin R & \mathit{if}\ u_{m,n} < \theta_{topic} \end{cases} \tag{4}$$

The segments in R are considered relevant and are further examined for inclusion in the expanding SAT set. As in SAT generation, the curator must select a search threshold that returns a number of search results that are useful and manageable.

3.5.2 Clustering search results

Using the method described above for SAT generation, search results are clustered and both clusters and individual segments are presented to the curator for evaluation. The curator selects segments that align with the meaning of the topic, which are added to the SAT set. Unselected segments are added to the rejected set. This process is repeated to expand the SAT set, while ignoring the segments in the expanding SAT set and the expanding rejected set, until no additional matching segments can be found.

3.5.3 Confirming completeness with n-gram search

Finally, we use an n-gram search in the review tool to complement the semantic similarity approach. The n-gram search provides a straightforward mechanism to locate specific terms or phrases. This allows the curator to search the corpus for key phrases appearing in accepted segments to ensure similar segments have not been missed in the corpus. At this point, the curator has the completed SAT set.

This procedure offers several key advantages. Firstly, it supports systematic refinement and validation of the candidate topic, improving the robustness and reliability of the expanded vocabulary. Secondly, it allows for the tracking of rejected segments, ensuring that only those truly reflecting the core of the topic are included. Lastly, by documenting curator decisions and segment selections, the process is transparent and reproducible.

3.6 SAT Review

A completed SAT set represents the final set of constitution segments that represent the candidate topic. A completed SAT set is evaluated by a panel of scholars who assess its relevance and accuracy in relation to the corpus. They consider its potential impact on the vocabulary structure and its contribution to the conceptual framework and decide whether to integrate the topic into the vocabulary. Although the topic should have undergone a thorough review before this stage, further revisions may still be necessary to ensure its substantive value. This may involve revising the segment set or reconsidering the topic's placement within the vocabulary structure. Throughout the process, no results are accepted at face value; rigorous scrutiny is maintained at every step, and all changes are automatically documented to ensure transparency.

3.7 SAT Integration

Once the final revisions are complete, the new topic is given a label and description before being formally integrated into the vocabulary. Every constitution segment belonging to the completed SAT set is automatically tagged with the new topic.

4. Application

We present results obtained by applying our methodology to create three new CCP topics summarized in Table 1.

4.1 SAT Generation

The SAT generation process involved iterative refinement, ensuring alignment of the topic formulations with the intended conceptual scope of the candidate topics. For example, our initial formulation of the *parents* topic—taken from the Amend Project that tracks proposed amendments to the U.S. Constitution (Lepore and Resch 2023)—noted it "Grants parents certain rights or duties regarding their children. May include the duty to provide for one's legal children, or the right to make decisions about their education." Testing this formulation against our corpus of national constitutions revealed a wider range of parental rights included in constitutions globally, including making decisions not

Table 1. New CCP topics created using the process described in this paper

Topic key	Topic label	Topic description
parents	Rights and duties of parents	Grants parents certain rights or duties regarding their children. May include the duty to provide for one's legal children, or the right to make decisions about their education, upbringing, and other aspects of their lives.
recall	Voter recall of elected officials	Recall elected officials and public servants. Gives the electorate the power to recall, remove from office, or revoke the mandate of authorities by popular vote. Recall procedures include citizen initiative, petition, motion.
replace	Constitutional replacement	Replacement, repeal, or total revision of the constitution. Enactment of a new constitution. Procedures for calling a constitutional convention or constituent assembly.

only about children's education but also about upbringing more broadly and many other aspects of their lives. We thus added these additional components to our formulation. Table 1 shows the final description of each topic, and Table 2 summarizes the outcomes of the SAT generation phase overall, including the resulting seed set sizes.

Table 2. SAT generation for three new CCP topics

Topic	Search threshold	Cluster threshold	Search results	SAT seed set
parents	0.63	0.71	203	81
recall	0.62	0.71	4,064	51
replace	0.63	0.71	3,308	19

Each candidate topic had unique advantages or challenges during this stage. For the *parents* topic, the relatively narrow linguistic range of the topic made SAT generation rather straightforward. The limited synonyms or alternative formulations available for parental rights and responsibilities (e.g., terms such as mother and father) resulted in a manageable and sharply defined set of possible matches to our new topic (203 constitution segments). This linguistic specificity facilitated straightforward identification of relevant segments, resulting in an accepted seed set containing 81 segments.

In contrast, the *recall* topic presented challenges due to both its conceptual specificity and the substantial volume of initial search results (4,064 segments). The term "recall" frequently appeared in diverse contexts involving multiple agents and procedures beyond the voter-initiated recall of elected officials, such as legislative- or executive-driven mechanisms. Given that our formulation explicitly targeted voter recall of elected officials, we encountered numerous false positives. To comprehensively address this issue, we selected a relatively lower similarity threshold to ensure extensive retrieval of potential matches, yielding 4,064 segments for review and ultimately 51 accepted segments. Although the lower similarity threshold required an extensive manual review to exclude irrelevant segments, it ultimately improved the precision and accuracy of the SAT seed set.

The *replace* topic involved conceptual overlaps, particularly with the existing CCP topic on constitutional amendment provisions. The semantic distinctions between "amendment" and "replacement" are often subtle and highly context-dependent, complicating the identification of relevant constitution segments. To mitigate this ambiguity, we iteratively refined our topic formulation, emphasizing explicit references to complete constitutional replacement or the enactment of entirely new constitutions. This refinement successfully clarified the conceptual boundary, resulting in a precise SAT seed set of 19 segments.

It is important to emphasize that SAT generation does not require identification of an exhaustive set of similar segments when selecting segments for the seed set. Relevant segments inadvertently overlooked at this stage are typically identified during the subsequent SAT expansion phase. This is because SAT expansion uses the seed set segments to search the corpus, and these initial seed set segments should exhibit higher semantic similarity to relevant segments than the original topic formulation used in SAT generation, ensuring that missed segments will likely emerge later. This iterative approach fosters continuous refinement, contributing to comprehensive and robust topic coverage.

4.2 SAT Expansion

SAT expansion begins with the seed set obtained during the SAT generation phase and iteratively searches for additional segments that align semantically with the topic. Tables 3, 4, and 5 summarize the SAT expansion results for the *parents*, *recall*, and *replace* topics. We adopted a deliberately conservative approach to inclusion throughout the SAT expansion phase, prioritizing conceptual precision over exhaustive inclusion. Segments were only accepted if they explicitly addressed the core topic—such as granting a specific right or imposing a specific duty—rather than merely implying relevance through indirect or contextual references. For example, this approach prompted us to accept the following constitutional provision in the *parents* SAT set since it speaks explicitly to parental rights and duties in noting that “[b]oth parents have the right and responsibility to raise their children.” It also prompted us to reject the following provision since it implies, but does not specifically impose, a duty on parents, noting “Children are equal in their rights regardless of their origin and whether they are born in or out of wedlock.” This cautious strategy helped preserve topic clarity and ensured the resulting SAT sets captured the most definitive and representative constitution segments related to those topics.

For the *parents* topic, the first SAT expansion round added 37 segments, increasing the total SAT size to 118 segments. Subsequent rounds yielded diminishing returns, ultimately stabilizing at 124 segments after five rounds of SAT expansion. The similarity threshold was lowered slightly after the third round to broaden the search scope, but this produced minimal further gains, indicating sufficient prior inclusion of relevant segments. Relevant segments reveal how constitutions typically frame parental rights alongside corresponding duties, particularly regarding children’s education and welfare. These provisions often reflect broader societal tensions between family autonomy and state protection interests, illustrating constitutional attempts to balance these competing values across different legal systems.

Table 3. SAT expansion for *parents* topic (Rights and duties of parents)

Round	Candidate Segments	Clusters	Segments Accepted	SAT Size
0[a]	203	12	81	81
1	292	19	37	118
2	71	8	5	123
3	12	3	0	123
4	194	17	1	124
5	15	1	0	124

A search threshold of 0.7 and a clustering threshold of 0.74 were used for rounds 1-4. A search threshold of 0.69 was used for rounds 5-6. The dashed line indicates the point at which the search threshold was reduced in order to increase the number of candidate segments.

[a] Round 0 is the SAT generation stage when the seed set is created (see Table 2).

The conceptual specificity of the *recall* topic required particular attention in this stage as well. After an initially large number of 1,549 matches to review, the rejections made in that round honed

Table 4. SAT expansion for *recall* topic (Voter recall of elected officials)

Round	Candidate Segments	Clusters	Segments Accepted	SAT Size
0[a]	4,064	39	51	51
1	1,549	24	10	61
2	253	6	5	66
3	92	9	3	69
4	43	5	0	69
5	1,639	84	2	71
6	58	4	3	74
7	20	2	1	75

A search threshold of 0.7 and a clustering threshold of 0.74 were used for rounds 1-4. A search threshold of 0.69 was used for rounds 5-7.

[a] Round 0 is the SAT generation stage when the seed set is created (see Table 2).

the topic focus and produced smaller segment sets for review in subsequent rounds, ultimately reaching a final SAT size of 75 segments. Notably, the threshold adjustment at the fifth round enabled the identification of additional relevant segments missed earlier, demonstrating the value of iterative threshold calibration. Many rejected segments referenced recall mechanisms initiated by legislatures or courts, underscoring the importance of human review of these matches to ensure constitution segments matched our topical focus on citizen-initiated recall. Accepted segments tended to come from a smaller group of constitutions—particularly those in Africa, Latin America, and East Asia—where provisions for popular recall are more common.

The *replace* topic experienced steady but limited growth during expansion rounds. Beginning with 19 segments in the seed set, it expanded to a final SAT size of 36 segments. The threshold adjustment midway through the expansion rounds allowed for the capture of one additional segments, though a further iteration yielded no new additions, confirming that comprehensive coverage had been achieved. One of the main challenges for this topic during expansion was distinguishing provisions about full constitutional replacement from those concerning major amendments or reform procedures. Accepted segments typically included clear references to constituent assemblies or the enactment of a new constitution, while more ambiguous cases were conservatively excluded. Notably, references to constitutional replacement were prevalent among Latin American constitutions.

Table 5. SAT expansion for *replace* topic (Constitutional replacement)

Round	Candidate Segments	Clusters	Segments Accepted	SAT Size
0[a]	3,308	15	19	19
1	1,448	21	9	28
2	321	18	3	31
3	336	16	4	35
4	49	6	0	35
5	1,474	94	1	36
6	125	7	0	36

A search threshold of 0.7 and a clustering threshold of 0.74 were used for rounds 1-4. A search threshold of 0.69 was used for rounds 5-6.

[a] Round 0 is the SAT generation stage when the seed set is created (see Table 2).

Clustering played a crucial role throughout the SAT expansion process by organizing search results into coherent groups of semantically related segments. This approach facilitated the identification of potential sub-topics and regional patterns within segments, enabling recognition of specific regional language patterns (which can be used to hone to topic formulations) or localized constitutional practices (which can inform our future analysis). For instance, clustering revealed that several constitutions from Latin America (e.g., Costa Rica, El Salvador, Panama, and Uruguay) explicitly stipulate that parents have identical obligations to children born out of wedlock as to those born within marriage, highlighting a clear regional pattern. Likewise, our clustering analysis identified a distinct cluster containing four segments from Sub-Saharan constitutions (Central African Republic, Gabon, Niger, Senegal) that uniquely references parents' entitlement to support from "public collectivities" in addition to the state.

Additionally, the n-gram search functionality in the review tool was an important complement to the semantic similarity approach. It provided a way to locate specific terms or phrases, enhancing precision when approving or rejecting candidate segments. This feature proved especially beneficial in identifying segments containing critical keywords. For example, segments related to *recall* frequently include the titular term itself or close synonyms such as "revoke" or "revocation." Conducting an n-gram search at the start or end of the validation process ensured curators systematically verified acceptance or rejection of all segments containing these key terms.

Overall, the SAT expansion results for these three topics demonstrate the effectiveness of our iterative methodology in systematically capturing a comprehensive set of relevant segments for each topic. By combining sentence-level semantic similarity, clustering, and n-gram search, the process enables not only the broad identification of conceptually aligned provisions but also the inductive discovery of sub-topics and regional patterns. Together, these features contribute to a transparent, reproducible, and scalable approach to vocabulary expansion, grounded in both linguistic precision and domain expertise.

4.3 SAT Review

The final SAT sets were presented to a panel of experts at the Comparative Constitutions Project for review and approval. The CCP research team currently consists of two directors, a research director, a research associate, and five senior research analysts. Upon review by our full research team, the *parents*, *recall*, and *replace* topics were accepted for inclusion into the CCP vocabulary.

4.4 SAT Integration

To conclude the process, the CCP corpus of national constitutions was automatically tagged with the new *parents*, *recall*, and *replace* topics. In other words, the topics were applied to the accepted SAT segments for each topic in the XML files that comprise our corpus of 192 in-force national constitutions. For example, the *parents* topic can now be viewed on the Constitute website. Using the SAT segments to tag the corresponding constitution sections thus avoids potential human errors of manual tagging such as failing to tag an accepted SAT segment, or erroneously tagging a segment that does not form part of the SAT. Additional new topics are in the process of being added to the CCP vocabulary with this method.

5. Discussion

The strength of our methodology lies in representing topics through sets of corpus segments—segments-as-topic (SAT)—rather than a single phrase that attempts to capture the meaning of a topic. Using corpus segments as topics provides better semantic similarity matching by capturing natural language patterns, contextual cues, and linguistic variations as they appear in the corpus. These segments implicitly represent different expressions of the same concept, which a single phrase cannot do.

Our methodology involves multiple stages leveraging both automated tools and human expertise to refine and validate our findings. Validation ultimately depends on human decisions, making false positives and false negatives less of a concern than in fully automated systems. Automated classification or tagging may incorrectly identify sections as matching a particular topic (false positives), or fail to identify segments that belong to the topic of interest (false negatives). In contrast, manually tagging sections relies entirely on human judgment, making it more susceptible to false negatives due to the difficulty of reading and accurately applying topics to every line of the world's constitutions.

The SAT method necessarily encourages the use of low search thresholds at the topic generation stage in order to harvest accepted *and* rejected results. Curators then identify and correct false positives by rejecting segments that do not align with the conceptual intent of the topic, or vice versa for false negatives. These results provide insight into performance, specifically whether a curator's formulated topic text is generating results with a satisfactory proportion of matching segments. Since a panel of scholars evaluates the final set, a form of inter-coder reliability is built into our process. If these domain experts conclude that some additional segment should be added to, or removed from, the final set defining a topic, the risk of false positives and false negatives is further mitigated.

We advocate combining automation with human expertise in constitutional design. Finding the proper balance, though, has been a learning process. Automated topic application without robust human review is unacceptable, for the reasons noted above. By contrast, our early efforts at comprehensive manual review were inefficient. Researchers had to sift through countless spreadsheets of machine-generated topic matches—only to start over whenever the topic formulations changed or the similarity threshold was adjusted. Through trial and error, we identified key steps that could be automated to strategically aid human review: initial topic matching, tracking accepted and rejected segments across iterations, clustering results, and documenting coding decisions. These components are now embedded in the SAT method and tool—now publicly available—which we see as a game-changing approach to integrating automation with human expertise in the expansion of conceptual vocabularies.

The selection of search and cluster thresholds in our methodology rests on empirical rather than theoretical foundations. Although sentence embeddings capture statistical patterns in language, they do not correspond to formal semantic representations and, thus, semantic similarity does not have a formal definition. Consequently, threshold selection becomes an inductive process determined by practical efficacy rather than theoretical imperatives. This flexibility allows curators to adjust thresholds during both SAT generation and expansion stages based on corpus-specific characteristics.

This empirical approach raises important questions about semantic similarity scores. Low-similarity pairings may not represent absence of semantic relationship but rather more subtle connections. Our analysis reveals that complex multi-concept sentences containing relevant sub-clauses often fall below arbitrary thresholds despite their topical relevance. While percentile-based alternatives present their own challenges (particularly in determining appropriate cutoffs), examining similarity score distributions can justify topic-specific thresholds tailored to each topic's unique distribution properties.

For clustering thresholds, we employ sensitivity analysis to evaluate the empirical stability of resulting structures. This data-driven approach reveals how clustering patterns respond to threshold adjustments, allowing us to characterize their robustness. Our analysis of cluster number plotted against threshold values reveals clear optimization points—thresholds that avoid both the over-clustering that occurs at low values and the fragmentation that occurs at high values. Additional metrics like cluster size distribution and singleton set size further inform threshold optimization. These investigations are guiding the development of an adaptive method for determining optimal cluster thresholds automatically.

To enhance comprehensiveness, we integrate n-gram search capabilities with our semantic similarity approach. This hybrid methodology addresses a key limitation: segments where the overall semantic similarity falls below threshold despite containing specific relevant phrases. This integration proves particularly valuable for long, multi-concept sentences where the relevant topic represents only a portion of the overall semantic content. By combining vector-based semantic similarity with targeted lexical searches, our methodology creates a more robust identification process that captures segments which might otherwise be missed by either approach in isolation. In other words, the SAT method may prove complementary to other techniques under certain circumstances.

6. Conclusion

The approach described here demonstrates how to expand a vocabulary by combining automated text classification and expert-driven topic curation. We have developed the segments-as-topic (SAT) methodology in which a topic is defined by a set of segments from a corpus. Using the SAT methodology we were able to identify and integrate new topics into the Comparative Constitutions Project (CCP) vocabulary. Importantly, all our team members have access to the software that implements the SAT methodology, giving us equal opportunity to propose new topics for collaborative expert review. By harnessing the individual initiatives of our domain experts, we ensure that our vocabulary remains up-to-date and reflects contemporary constitutional discourse.

The methodology also has considerable potential for tasks and domains beyond CCP. For example, lawyers and legal researchers often sift through vast amounts of case law to find relevant precedents and legal principles. The SAT methodology could be used to automate the classification and integration of new case law into existing legal taxonomies, making it easier to identify pertinent sections of case law when conducting research.

Within CCP, the methodology will be used to expand the range of topics CCP tracks in national constitutions to incorporate new topics being added to constitutions in recent years. It could also be utilized to create specialized sub-vocabularies of topics. For example, a sub-vocabulary related to constitution reform and drafting would enable researchers to track the evolution of themes across public consultation responses, the deliberations of drafting bodies, and versions of constitutional texts.

The methodology not only serves our forward-looking objectives discussed above, but also encompasses retrospective goals. Most importantly, our next step is to expand the application of existing topics in the CCP vocabulary that were formulated before we adopted semantic similarity technology. In the past, these topics were manually tagged by the CCP team, meaning that we searched through our corpus of 192 constitutions, as well as a number of historical and draft constitutional texts, for specific provisions that matched the corresponding topics. Using the SAT approach, we can now identify additional constitutional provisions that may have been overlooked in our manual tagging process. In other words, our methodology enables us to reduce the margin of human error and ensure a more comprehensive exploration of our corpus, maximizing the coverage and depth of our topic integration efforts. In essence, we are not just expanding our vocabulary; we are expanding our understanding of the founding ideals people value most around the world.

Acknowledgments We would like to thank Jessie Baugher, Lucas Elkins, Kendall Lowe, Maya Mackey, Jamie Mahowald, Catalina Mulhollan, Elias Roldan, and Yaser Tahboub for their research assistance at the inception of this project.

Funding Statement This work was supported by the National Science Foundation award number 2315189, awarded to Z.E. and A.M. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests The authors have declared that no competing interests exist.

Data Availability Statement All data used in this study as well as replication materials are available in a repository on Zenodo:

Ethical Standards The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Author Contributions Conceptualization: R.G.; M.M.; A.C.; Z.E.; A.M.; G.P. Data curation: R.G. Formal analysis: R.G.; M.M. Funding acquisition: Z.E.; A.M. Methodology: R.G. Project administration: A.M. Software: R.G. Supervision: Validation: R.G.; M.M.; A.M. Visualization: M.M. Writing original draft: R.G.; M.M. Writing review and editing: R.G.; M.M.; A.C.; Z.E.; A.M.; G.P. All authors approved the final submitted draft.

Notes

References

Aggarwal, Charu C. 2022. *Machine learning for text.* Second Edition. Springer.

Bestvater, Samuel E., and Burt L. Monroe. 2023. Sentiment is Not Stance: Target-Aware Opinion Classification for Political Text Analysis. *Political Analysis* 31, no. 2 (April): 235–256. ISSN: 1047-1987, 1476-4989, accessed April 24, 2025. https://doi.org/10.1017/pan.2022.10. https://www.cambridge.org/core/journals/political-analysis/article/sentiment-is-not-stance-targetaware-opinion-classification-for-political-text-analysis/743A9DD62DF3F2F448E199BDD1C37C8D.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3 (Jan): 993–1022.

Cer, Daniel, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, et al. 2018. *Universal sentence encoder.* arXiv: 1803.11175 [cs.CL].

Collier, David, and James E. Mahon. 1993. Conceptual "Stretching" Revisited: Adapting Categories in Comparative Analysis. *The American Political Science Review* 87 (4): 845–855.

Cruz, Andrés, Zachary Elkins, Roy Gardner, Matthew Martin, and Ashley Moran. 2023. Measuring constitutional preferences: A new method for analyzing public consultation data. Edited by Jerg Gutmann. *PLOS ONE* 18, no. 12 (December). ISSN: 1932-6203. https://doi.org/10.1371/journal.pone.0295396.

Dieng, Adji B, Francisco JR Ruiz, and David M Blei. 2020. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics* 8:439–453.

Elkins, Zachary, and Tom Ginsburg. 2025 [2005]. *Comparative Constitutions Project.* https://comparativeconstitutionsproject.org/.

Eshima, Shusei, Kosuke Imai, and Tomoya Sasaki. 2024. Keyword-assisted topic models. *American Journal of Political Science* 68 (2): 730–750.

Farouk, Mamdouh. 2018. Sentence Semantic Similarity based on Word Embedding and WordNet. In *2018 13th International Conference on Computer Engineering and Systems (ICCES),* 33–37. December. https://doi.org/10.1109/ICCES.2018.8639211.

Gardner, Roy. 2023. Semantic Analysis to Support Peace Analytics. *Peace Analytics Series. PeaceRep: The Peace and Conflict Resolution Evidence Platform, University of Edinburgh,* https://peacerep.org/publication/semantic-analysis-to-support-peace-analytics/.

Gerring, John. 2011. *Social science methodology: a unified framework.* 2nd ed. Strategies for Social Inquiry. Cambridge University Press.

Goertz, Gary. 2020. *Social Science Concepts and Measurement.* New and completely revised edition. Princeton Oxford: Princeton University Press. ISBN: 978-0-691-20548-9.

Grootendorst, Maarten. 2022. Bertopic: neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794.*

Jones, Bryan D, Frank R Baumgartner, Sean M Theriault, Derek A Epp, Cheyenne Lee, Miranda E Sullivan, and Chris Cassella. 2023. Policy agendas project: codebook. *Comparative Agendas Project. URL: https://www. comparativeagendas.net/us.*

Lehmann, Pola, Simon Franzmann, Tobias Burst, Sven Regel, Felicia Riethmüller, Andrea Volkens, Bernhard Weßels, and Lisa Zehnter. 2022. The manifesto data collection. *Manifesto Project (MRG/CMP/MARPOR). Version 2022a.*

Lepore, Jill, and Tobias Resch. 2023. *Amendments Project.* https://amendmentsproject.org/.

Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Edoardo M Airoldi, et al. 2013. The structural topic model and applied social science. In *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation,* 4:1–20. 1. Harrahs and Harveys, Lake Tahoe.

Rodriguez, Pedro L, and Arthur Spirling. 2022. Word embeddings: what works, what doesn't, and how to tell the difference for applied research. *The Journal of Politics* 84 (1): 101–115.

Sartori, Giovanni, ed. 1984. *Social Science Concepts: A Systematic Analysis.* 1st ed. Social Science Concepts. Beverly Hills, CA: Sage Publications. ISBN: 978-0-8039-2177-1.

Whorf, Benjamin Lee, John Bissell Carroll, Stephen C. Levinson, and Penny Lee. 2012 [1956]. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf* [in eng]. 2nd ed. Cambridge (Mass.): the MIT press. ISBN: 978-0-262-51775-1.

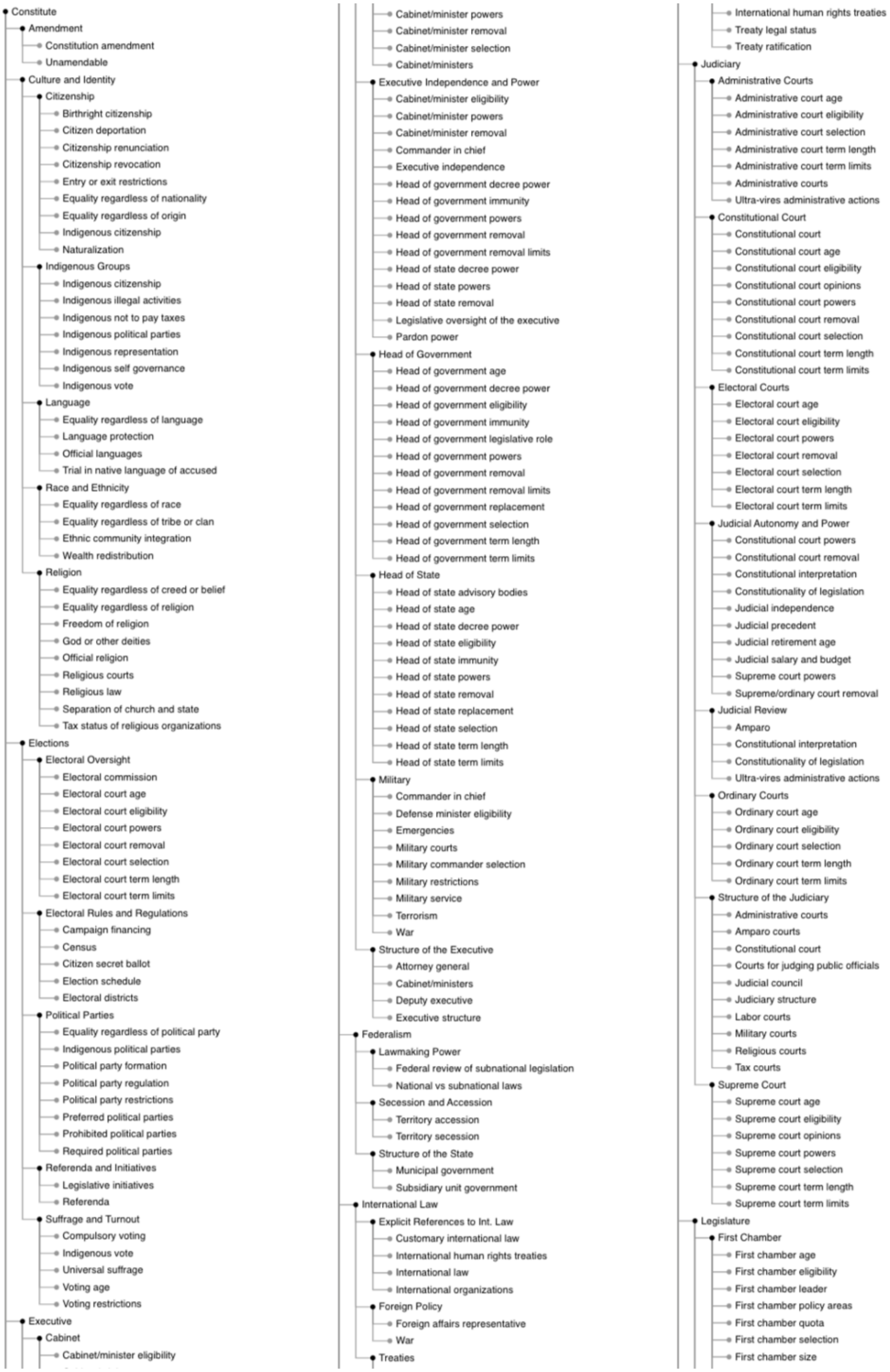
Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus.* Translated by Frank P. Ramsey and Charles Kay Ogden. 573–588. London, United Kingdom: Kegan Paul, Trench, Trubner & Co. Ltd., May.

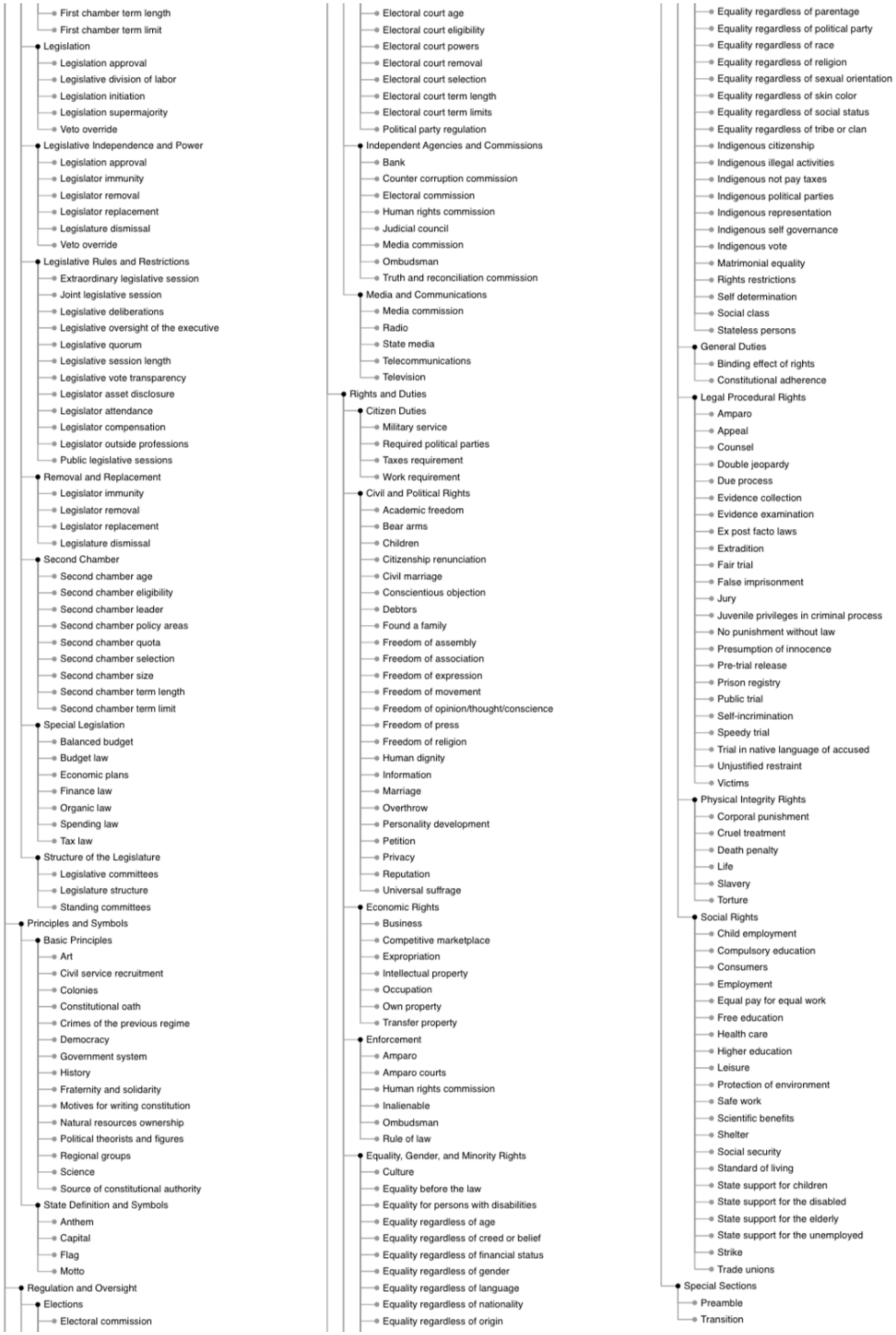
Yang, Mingming, Rui Wang, Kehai Chen, Masao Utiyama, Eiichiro Sumita, Min Zhang, and Tiejun Zhao. 2019. Sentence-level agreement for neural machine translation. In *Proceedings of the 57th annual meeting of the association for computational linguistics,* edited by Anna Korhonen, David Traum, and Lluís Màrquez, 3076–3082. Florence, Italy: Association for Computational Linguistics, July. https://doi.org/10.18653/v1/P19-1296. https://aclanthology.org/P19-1296.

Ying, Luwei, Jacob M Montgomery, and Brandon M Stewart. 2022. Topics, concepts, and measurement: a crowdsourced procedure for validating topics as measures. *Political Analysis* 30 (4): 570–589.

Appendix 1. CCP Vocabulary

Below is a representation of the 334 topics in the CCP ontology, organized into 12 categories.


Constitute
Amendment
Constitution amendment
Unamendable
Culture and Identity
Citizenship
Birthright citizenship
Citizen deportation
Citizenship renunciation
Citizenship revocation
Entry or exit restrictions
Equality regardless of nationality
Equality regardless of origin
Indigenous citizenship
Naturalization
Indigenous Groups
Indigenous citizenship
Indigenous illegal activities
Indigenous not to pay taxes
Indigenous political parties
Indigenous representation
Indigenous self governance
Indigenous vote
Language
Equality regardless of language
Language protection
Official languages
Trial in native language of accused
Race and Ethnicity
Equality regardless of race
Equality regardless of tribe or clan
Ethnic community integration
Wealth redistribution
Religion
Equality regardless of creed or belief
Equality regardless of religion
Freedom of religion
God or other deities
Official religion
Religious courts
Religious law
Separation of church and state
Tax status of religious organizations
Elections
Electoral Oversight
Electoral commission
Electoral court age
Electoral court eligibility
Electoral court powers
Electoral court removal
Electoral court selection
Electoral court term length
Electoral court term limits
Electoral Rules and Regulations
Campaign financing
Census
Citizen secret ballot
Election schedule
Electoral districts
Political Parties
Equality regardless of political party
Indigenous political parties
Political party formation
Political party regulation
Political party restrictions
Preferred political parties
Prohibited political parties
Required political parties
Referenda and Initiatives
Legislative initiatives
Referenda
Suffrage and Turnout
Compulsory voting
Indigenous vote
Universal suffrage
Voting age
Voting restrictions
Executive
Cabinet
Cabinet/minister eligibility
Cabinet/minister powers
Cabinet/minister removal
Cabinet/minister selection
Cabinet/ministers
Executive Independence and Power
Cabinet/minister eligibility
Cabinet/minister powers
Cabinet/minister removal
Commander in chief
Executive independence
Head of government decree power
Head of government immunity
Head of government powers
Head of government removal
Head of government removal limits
Head of state decree power
Head of state powers
Head of state removal
Legislative oversight of the executive
Pardon power
Head of Government
Head of government age
Head of government decree power
Head of government eligibility
Head of government immunity
Head of government legislative role
Head of government powers
Head of government removal
Head of government removal limits
Head of government replacement
Head of government selection
Head of government term length
Head of government term limits
Head of State
Head of state advisory bodies
Head of state age
Head of state decree power
Head of state eligibility
Head of state immunity
Head of state powers
Head of state removal
Head of state replacement
Head of state selection
Head of state term length
Head of state term limits
Military
Commander in chief
Defense minister eligibility
Emergencies
Military courts
Military commander selection
Military restrictions
Military service
Terrorism
War
Structure of the Executive
Attorney general
Cabinet/ministers
Deputy executive
Executive structure
Federalism
Lawmaking Power
Federal review of subnational legislation
National vs subnational laws
Secession and Accession
Territory accession
Territory secession
Structure of the State
Municipal government
Subsidiary unit government
International Law
Explicit References to Int. Law
Customary international law
International human rights treaties
International law
International organizations
Foreign Policy
Foreign affairs representative
War
Treaties
International human rights treaties
Treaty legal status
Treaty ratification
Judiciary
Administrative Courts
Administrative court age
Administrative court eligibility
Administrative court selection
Administrative court term length
Administrative court term limits
Administrative courts
Ultra-vires administrative actions
Constitutional Court
Constitutional court
Constitutional court age
Constitutional court eligibility
Constitutional court opinions
Constitutional court powers
Constitutional court removal
Constitutional court selection
Constitutional court term length
Constitutional court term limits
Electoral Courts
Electoral court age
Electoral court eligibility
Electoral court powers
Electoral court removal
Electoral court selection
Electoral court term length
Electoral court term limits
Judicial Autonomy and Power
Constitutional court powers
Constitutional court removal
Constitutional interpretation
Constitutionality of legislation
Judicial independence
Judicial precedent
Judicial retirement age
Judicial salary and budget
Supreme court powers
Supreme/ordinary court removal
Judicial Review
Amparo
Constitutional interpretation
Constitutionality of legislation
Ultra-vires administrative actions
Ordinary Courts
Ordinary court age
Ordinary court eligibility
Ordinary court selection
Ordinary court term length
Ordinary court term limits
Structure of the Judiciary
Administrative courts
Amparo courts
Constitutional court
Courts for judging public officials
Judicial council
Judiciary structure
Labor courts
Military courts
Religious courts
Tax courts
Supreme Court
Supreme court age
Supreme court eligibility
Supreme court opinions
Supreme court powers
Supreme court selection
Supreme court term length
Supreme court term limits
Legislature
First Chamber
First chamber age
First chamber eligibility
First chamber leader
First chamber policy areas
First chamber quota
First chamber selection
First chamber size


First chamber term length
First chamber term limit
Legislation
Legislation approval
Legislative division of labor
Legislation initiation
Legislation supermajority
Veto override
Legislative Independence and Power
Legislation approval
Legislator immunity
Legislator removal
Legislator replacement
Legislature dismissal
Veto override
Legislative Rules and Restrictions
Extraordinary legislative session
Joint legislative session
Legislative deliberations
Legislative oversight of the executive
Legislative quorum
Legislative session length
Legislative vote transparency
Legislator asset disclosure
Legislator attendance
Legislator compensation
Legislator outside professions
Public legislative sessions
Removal and Replacement
Legislator immunity
Legislator removal
Legislator replacement
Legislature dismissal
Second Chamber
Second chamber age
Second chamber eligibility
Second chamber leader
Second chamber policy areas
Second chamber quota
Second chamber selection
Second chamber size
Second chamber term length
Second chamber term limit
Special Legislation
Balanced budget
Budget law
Economic plans
Finance law
Organic law
Spending law
Tax law
Structure of the Legislature
Legislative committees
Legislature structure
Standing committees
Principles and Symbols
Basic Principles
Art
Civil service recruitment
Colonies
Constitutional oath
Crimes of the previous regime
Democracy
Government system
History
Fraternity and solidarity
Motives for writing constitution
Natural resources ownership
Political theorists and figures
Regional groups
Science
Source of constitutional authority
State Definition and Symbols
Anthem
Capital
Flag
Motto
Regulation and Oversight
Elections
Electoral commission
Electoral court age
Electoral court eligibility
Electoral court powers
Electoral court removal
Electoral court selection
Electoral court term length
Electoral court term limits
Political party regulation
Independent Agencies and Commissions
Bank
Counter corruption commission
Electoral commission
Human rights commission
Judicial council
Media commission
Ombudsman
Truth and reconciliation commission
Media and Communications
Media commission
Radio
State media
Telecommunications
Television
Rights and Duties
Citizen Duties
Military service
Required political parties
Taxes requirement
Work requirement
Civil and Political Rights
Academic freedom
Bear arms
Children
Citizenship renunciation
Civil marriage
Conscientious objection
Debtors
Found a family
Freedom of assembly
Freedom of association
Freedom of expression
Freedom of movement
Freedom of opinion/thought/conscience
Freedom of press
Freedom of religion
Human dignity
Information
Marriage
Overthrow
Personality development
Petition
Privacy
Reputation
Universal suffrage
Economic Rights
Business
Competitive marketplace
Expropriation
Intellectual property
Occupation
Own property
Transfer property
Enforcement
Amparo
Amparo courts
Human rights commission
Inalienable
Ombudsman
Rule of law
Equality, Gender, and Minority Rights
Culture
Equality before the law
Equality for persons with disabilities
Equality regardless of age
Equality regardless of creed or belief
Equality regardless of financial status
Equality regardless of gender
Equality regardless of language
Equality regardless of nationality
Equality regardless of origin
Equality regardless of parentage
Equality regardless of political party
Equality regardless of race
Equality regardless of religion
Equality regardless of sexual orientation
Equality regardless of skin color
Equality regardless of social status
Equality regardless of tribe or clan
Indigenous citizenship
Indigenous illegal activities
Indigenous not pay taxes
Indigenous political parties
Indigenous representation
Indigenous self governance
Indigenous vote
Matrimonial equality
Rights restrictions
Self determination
Social class
Stateless persons
General Duties
Binding effect of rights
Constitutional adherence
Legal Procedural Rights
Amparo
Appeal
Counsel
Double jeopardy
Due process
Evidence collection
Evidence examination
Ex post facto laws
Extradition
Fair trial
False imprisonment
Jury
Juvenile privileges in criminal process
No punishment without law
Presumption of innocence
Pre-trial release
Prison registry
Public trial
Self-incrimination
Speedy trial
Trial in native language of accused
Unjustified restraint
Victims
Physical Integrity Rights
Corporal punishment
Cruel treatment
Death penalty
Life
Slavery
Torture
Social Rights
Child employment
Compulsory education
Consumers
Employment
Equal pay for equal work
Free education
Health care
Higher education
Leisure
Protection of environment
Safe work
Scientific benefits
Shelter
Social security
Standard of living
State support for children
State support for the disabled
State support for the elderly
State support for the unemployed
Strike
Trade unions
Special Sections
Preamble
Transition