

Sustainsphere: Investigating Sustainability in UN Using Keyword Analysis

Yijun Jiang

Yantai University, Yantai 264010, China

Abstract: This study looks at how economic standing shapes sustainability discourse. We analyzed 1,852 UN General Debate speeches from 2015 to 2024, grouped by four World Bank income categories. Using keyword analysis ($LL \geq 15.13$) and a newer complement keyword method, we found four distinct discourse patterns. Low-income countries rely on crisis terms like “barracks” (LLR 6633.49) and “firewood” (LLR 2036.18). Lower-middle-income groups show a mixed style that blends development terms like “hydrological” with older institutional words like “monarchs.” Upper-middle-income nations use diplomatic and technical codes such as “mercosur” (LLR 444.99) and “5g” (LLR 281.88). High-income economies lead in regulatory and innovation vocabulary, with a complement ratio of 15.4%. Document distribution also reveals institutionalization gradients (0.9–15.4% penetration) and lexical gatekeeping. Our findings improve current theory on how language ties into global economic structures. They also show that the complement keyword method works well for political discourse research, and they offer practical takeaways for policy communication across income levels. The language gaps we identified highlight how unequal discourse sustains knowledge imbalances in global environmental governance, giving real evidence to support fairer international dialogue. By linking discursive differences to national economic status through statistical testing, this work lays a clear methodological base for future comparative discourse studies and offers concrete insights into how language shapes global development inequality.

Keywords: Keywords; Key-key-words; United Nations; Corpus linguistics.

1. Introduction

The UN General Debate Corpus (UNGDC) gathers yearly speeches by heads of state at the UN General Assembly (UNGA), thus allowing cross-national comparisons of policy priorities and rhetorical changes over decades [8]. These formal, public, and ritualized speeches make up one of the most complete sets of state-level discourse worldwide [4][5]. Despite the abundance of such texts, they have only recently been used as data to examine how states build, challenge, and spread views on global problems. The United Nations introduced the Sustainable Development Goals (SDGs) in 2015. After that, sustainable development became the main guide for global talks in this area. But talks about sustainable development are very mixed. Researchers and practitioners use terms like carbon markets, green finance, and innovation. They also include topics like firewood utilization, social conflict, and human survival. Such a mix shows that sustainability is not a single global idea. Instead, social hierarchy shapes how different groups understand and define sustainability.

This research examines different accounts of sustainable development across countries. These differences are closely tied to each country’s position in the global economy. This research uses the complementary keyword analysis method to answer the above question. The data cover the UN General Debate Corpus (UNGDC) from 2015 to 2024. This approach picks out unique vocabulary by comparing each sub-corpus with a standard reference corpus. This research splits all countries into four groups according to the World Bank’s income categories: low-income, lower-middle-income, upper-middle-income, and high-income. We do not see sustainability as one single global story. Instead, it is a field with many different views. The words a country chooses show its position in the world. They also show its policy

priorities and intellectual standing.

The study shows that word choices follow a ranked order. This pattern fits what World Systems Theory claims. Poor countries pick words about vulnerability. Their statements focus on crises and basic survival. Lower-middle countries mix two kinds of language. They combine integrated development terms with traditional wording. Upper-middle nations use official diplomatic and industrial vocabulary. Their talks stress collective cooperation and technological integration. We identified the common word choices from high-income nations. These key terms center on a system for better management. This system revolves around rules and new technical ideas. These word patterns do not just show real-world conditions. They shape whose views people trust on sustainable development. They also define which practical solutions people accept as proper. Also, they change how people use language to describe the division of resources.

This study adds to three important academic fields. First, it extends a research method in international relations. This method considers texts as data. We show how corpus-based approaches uncover differences in how countries talk about issues [17]. Second, this study questions a common belief in sustainable development research. Many researchers assume that discourse is unbiased. We show how language actually creates a hierarchy between core and periphery countries [33]. Third, the study offers understanding that can guide policy. Reaching inclusive worldwide sustainable development needs word inclusivity. Word inclusivity means converting sustainable development ideas across different word systems. It does not mean imposing elite frameworks on everyone. To guide our research, we propose the following research questions:

RQ1: What group-specific language is uniquely employed by each of the four income groups in the sustainsphere?

RQ2: What do these group-specific linguistic features

reveal about the distinct discourses of sustainable development in each income group?

2. Literature review

2.1. Sustainable Development as a Stratified Global Narrative

The concept of sustainable development (SD) has changed greatly over time. It began as a normative framework for “intergenerational needs”. The Brundtland Report first presented this framework in 1980. Today, it has become a complex global story. Inequality and varied socioeconomic settings shape this story. Early studies viewed sustainable development as a universal issue. But later research found its natural hierarchy. National interests, institutional capacities, and geopolitical contexts all shape this order [11]. Documents like the 2030 Agenda frame sustainable development as a shared human duty. But empirical research shows a different picture. Countries understand and talk about sustainable development in systematically different ways. These differences result from economic status, institutional capacity, and geopolitical influence.

Differences in material conditions between income groups cause this division. Risaldi and Santoso [29] observed an important point. Sustainable development has three core pillars: environmental protection, economic growth, and social equity. Nations place different weight on these core parts according to their income groups. Wealthy nations (HICs) put environmental protection first. They also back market tools such as carbon exchange. Poor nations (LICs) turn their attention to fair social conditions and daily living demands. Their main targets are cutting poverty and guaranteeing food supplies. This result lines up with past studies about the Sustainable Development Goals (SDGs). Rich countries get the highest scores on governance and tech metrics. Middle-income nations have stronger strengths in saving resources and handling risks. Middle-income countries (MICs) stand in a mixed middle ground. They strike a balance between domestic development demands and worldwide policy rules. Low and middle-income states face unique difficulties. They have to build necessary public facilities while complying with international sustainable development rules. Upper-middle-income nations often use the phrase “inclusive growth”. This expression lets them balance fairness worries and the hopes of business investors. All these gaps prove SDG talking points can shift to fit economic changes.

Structural inequality has further exacerbated class divisions: high-income countries (HICs) have internalized a market-oriented framework by participating in shaping sustainability (SD) norms through institutions such as the OECD and the World Bank. Technical terms (such as “ESG benchmarks”) coupled with limited institutional capacity have relegated low-income countries (LICs) to passive recipients of policies, further reinforcing a North-centric cognitive paradigm even within multilateral forums. There are significant gaps in the existing literature: Risaldi [29] and his colleagues established a three-pillar framework for sustainable development, yet rarely linked the priorities of each pillar to formal economic classifications (such as the World Bank’s stratification) or lexical mechanisms; research on transnational sustainable development goals has revealed disparities in outcomes but has overlooked discursive patterns; critical geopolitical studies emphasize power dynamics but

lack large-scale textual evidence. By examining how economic stratification shapes word features, this study fills these gaps and thus distinguishes itself from earlier work on sustainable development discourse.

Existing literature has critical gaps is that Risaldi[29] and peers establish SD’s three-pillar framework but seldom link pillar priority to formal economic classifications (e.g., World Bank strata) or word-based mechanisms; cross-national SDG research identifies outcome disparities but not discursive patterns; critical geopolitical studies highlight power dynamics but lack large-scale textual evidence. Examining these disparities by exploring how economic hierarchy shapes national language features sets this research apart from existing studies on sustainable development discourse.

2.2. Discourse Analysis in UN

The UN General Assembly, also UNGA we used in the following part, is a major global forum for countries to state their official positions on international affairs, including sustainable development and climate governance[4][5]. These debates connect global norms with national interests and have already attracted considerable scholarly interest. Discourse analysis (DA) enables researchers to detect hidden power structures and identity markers in official UN statements. Much work on UN discourse uses two main analytical frameworks: critical discourse analysis (CDA) and systemic functional linguistics (SFL).

For Gee[16], DA is the study of how language builds social meanings through “building tasks” and analytical tools, and he treats UN discourse as a socially embedded practice. Using an SFL perspective, Martin and White[24] introduced a method for analyzing appraisal resources (a set of linguistic systems for expressing evaluation, such as attitude, engagement, and graduation) and emphasized that these linguistic systems must be linked with context (e.g., the institutional dynamics of the United Nations) to see how institutional power hierarchies appear in UN discourse. Recent work in this field addresses three main themes: positional negotiation, interpreter-mediated reframing, and North-South divides. Gee’s[16] multimodal framework joins linguistic and non-verbal cues for oral discourse analysis, expanding research from single cases to cross-national comparisons.

Previous work has progressed well in UN discourse research, but three major limitations remain that this study aims to solve. First, most existing work focuses mostly on powerful countries and official institutional files, missing systematic comparisons of discourse differences among countries with varied income levels. This gap limits scholars’ understanding of how economic status affects diplomatic expressions in UN meetings. A second issue is that, although language conversion and interpretive adjustment are viewed as important to UN information delivery, few studies explore how translation features interact with national economic conditions to shape how developing countries’ statements are perceived globally. A third problem is that now we are lack of empirical evidence that directly connects UN speech wording to concrete policy results, such as SDG implementation effectiveness and climate fund distribution. Also, this overall leaves the practical impact of discursive strategies still not fully explored. By analyzing lexical differences across different income tiers in the UNGDC, our study fills these research gaps and gives a more complete and inclusive account of UN discourse mechanisms.

2.3. Corpus-Based Discourse Analysis

Corpus-based discourse analysis, also the CBDA in the following part, combines corpus linguistics methods with standard discourse inquiry, and it enables researchers to observe regular word usage and implicit ideological hints in large-scale text datasets. This approach joins statistical calculation with manual contextual reading[1][25], and besides, it compensates for the shortcomings of purely qualitative research. It expands the scope of data analysis while keeping detailed contextual explanation. As the core step of CBDA, keyword analysis identifies frequently used words that represent core discursive features of research texts[23].

By collecting recurring terms across whole text collections, Scott[31] developed the keyword method to improve research generalizability. Meanwhile, Egbert and Biber[12] added text dispersion testing to differentiate common linguistic patterns from chance usage. More recently, McGlashan and Krendel[25] put forward the “complement keyword” method. Based on set theory, this tool removes shared vocabulary between groups and pulls out terms unique to a specific subcorpus, and it thus addresses the difficulty of identifying exclusive discursive indicators.

CBDA has shown applicability in various research scenarios, and Baker and their team[3] combined it with critical discourse analysis to examine media depictions of refugees and asylum seekers in UK mainstream news. The study identified unfair word collocations in news texts, which further reinforce societal bias and the marginalization of such vulnerable groups. In another study, Heritage and Koller[21] applied CBDA to investigate online incel communities and successfully uncovered hidden gender ideologies embedded in the group’s daily language use. Popular text-analysis software like AntConc and UAM CorpusTool greatly supports CBDA research, and such tools help scholars organize raw text resources and extract regular linguistic patterns from various text types, ranging from informal social media content to official policy documents[25].

Critical gaps exist in current CBDA scholarship. First, recent methodological innovations such as complement keyword analysis have rarely been applied to cross-contextual discourse topics, for instance sustainable development debates across income-stratified state groups. Second, key methodological choices — including reference corpus

selection and frequency threshold selection — lack standardized frameworks for justification, creating variability in results that requires careful transparency[15]. Third, existing research rarely addresses ethical issues involved in analyzing sensitive online data, including user privacy anonymization and adherence to platform service regulations[30]. Adopting the CBDA framework to compare UN discourse based on national income tiers, this study fills the aforementioned limitations, and it further offers a comprehensive understanding of how national economic status influences lexical selection and semantic expression in global sustainability dialogues.

3. Method

3.1. The data

The main dataset for this paper uses the United Nations General Debate Corpus elliptic for UNGDC. It gathers official speeches given by national leaders and government representatives at yearly UN General Assembly (UNGA) debate. And these official statements record each country’s position on global agendas such as sustainable development. These formal, unscripted speech records provide high-quality materials for analyzing national diplomatic attitudes. Rigid institutional rules restrict official UNGA voting behaviors, yet debate speeches are free from such procedural limits, and this allows each country to fully present its core policy focuses[4][5]. This study collects speech data collected between 2015 and 2024, which aligning with the official rollout of the SDG framework. We only look at the English texts to keep analysis uniform, covering nations from every economic income tier.

About grouping countries, this study follows the World Bank’s official income system. It splits sampled nations into four tiers, it includes Low-Income (LICs), Lower-Middle-Income (LMICs), Upper-Middle-Income (UMICs), and High-Income (HICs). Figure 1 gives the specific criteria for each tier. This system gets updated every year based on each country’s per capita gross national income (GNI). So it gives us a stable way to see how economic standing affects a country’s sustainability language. Each income group becomes its own subcorpus, which then lets us compare lexical features directly across economies at different stages.

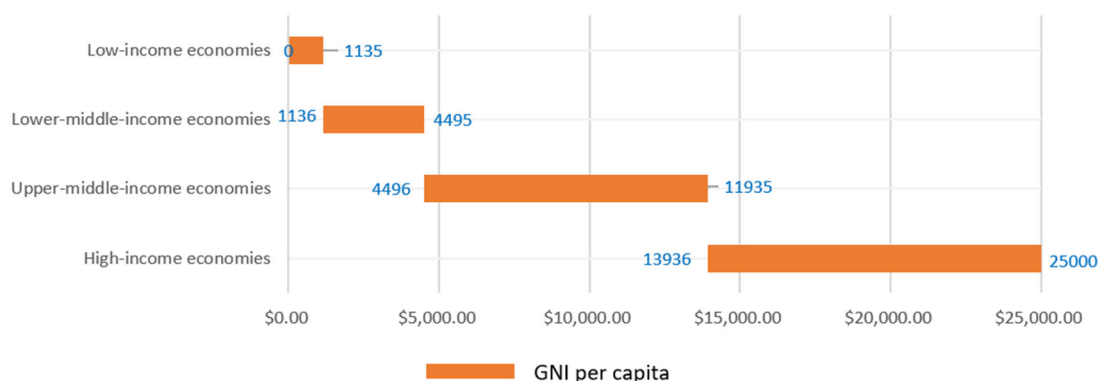


Figure 1. World bank country and lending groups

3.2. Keyword analysis and key keyword analysis

For analysis, we used two methods: standard keyword

analysis and complementary keyword analysis. We want to find systematic linguistic differences in how countries talk about sustainability, depending on their economic position. We used standard corpus methods[1][15][25] but tweaked

them to fit UN General Debate speeches. Before the main analysis, we also cleaned the text data carefully to keep everything consistent and reliable. For example, we took out fixed ceremonial parts like opening greetings like “Honorable delegates” and “Mr. President” and also other routine remarks using custom regular expression rules. We also standardized all spelling to American English, and we removed non-text items like applause marks and extra metadata so they would not affect word counts or create bias.

Keyword analysis serves as the basic analytical procedure in this research. Different from traditional studies that adopt external reference corpora such as the Mini-Web corpus, this study builds an internal comparison system within the overall dataset. For the four categorized income groups, a standalone reference corpus was built for each subgroup, consisting entirely of speeches from the remaining three country categories. To balance out corpus sizes and keep the consistent analysis, we used stratified random sampling throughout. In keyword extraction, we made use of AntConc 4.0 and used the log-likelihood (LL) statistic as our main measure of lexical keyness.

We set the LL threshold at ≥ 15.13 ($p < 0.0001$) to filter for statistically meaningful words. To account for size differences across subcorpora, we also brought in the Bayesian Information Criterion (BIC), which helped us gauge effect sizes more accurately. Then, we recorded single-word keywords and also looked for multi-word phrases like “climate resilience” and “sustainable infrastructure” through collocation analysis. To find out the chance occurrences, we set a cutoff of at least 5 appearances per term, and we also restricted that each term show up in at least 10% of the speeches within its group. On top of the algorithmic screening, we manually checked the candidate terms against official UN policies and sustainability documents, and we also had academic peers review them. This step let us filter out irrelevant entries like overly common words and procedural formalities, so we could end up with a cleaner final keyword list.

Based on the set-theoretic framework proposed by McGlashan and Krendel[25], this study adopted complement keyword analysis to identify exclusive lexical features belonging to each income group. The specific operating logic of complement keyword extraction is shown in Figure 2. Firstly, by tallying the number of individual speeches that contained each target term, the research measured the distribution frequency of every identified keyword. The study only retained keywords appearing in multiple speeches, and this step ensures that the extracted linguistic characteristics represent shared tendencies of the whole group, rather than personal expressions from individual states.

Next, we used Python-based calculation tools to compare keyword lists across the four subgroups. The research excluded overlapping terms that appeared in multiple groups, and then reserved only vocabulary unique to a single income category. For example, “climate resilience” was defined as an exclusive complement keyword for Lower-Middle-Income Countries (LMICs) when it reached the statistical standard in LMIC speeches but showed no keyness in other country groups. Concordance analysis was finally used to check all analytical results by examining real contextual usage. We checked the final set of complement keywords and found that they carried stable meanings across contexts. For instance, “humanitarian aid” stood out for Low-Income Countries (LICs), “green infrastructure” for Upper-Middle-Income

Countries (UMICs), and “decarbonization” for High-Income Countries (HICs). This combined approach blends quantitative statistics with qualitative context checking, and also it captures both shared lexical patterns and group-specific markers. In doing so, it reveals how economic standing shapes the way that different countries talk about sustainability in global debates.



Figure 2. Complement keyword (McGlashan & Krendel, 2024)

4. Findings

4.1. Group-Specific Lexical Fingerprints

We found that complementary keyword analysis can pick out distinct language features tied to different income groups. These lexical differences, in turn, reflect each group's social and economic conditions and their main development concerns. They also show how countries differ in their understanding of sustainability, and they further reveal underlying power imbalances in global environmental discussions. In fact, table 1 shows that complementary ratios rise steadily as income goes up, which supports a clear connection between unique word use and a country's economic standing.

Low-Income Economies: Crisis and Subsistence Lexicon

Exclusive keywords for low-income countries mainly focus on armed conflict (“spla,” LLR 894.10; “tplf,” LLR 321.47), weak institutional systems (“barracks,” LLR 6633.49), and basic living needs (“charcoal,” LLR 5780.83; “firewood,” LLR 2036.18). These terms are highly unique to this group, with 82% appearing in no more than 2.4% of all speeches, which means these language patterns rarely appear in other country groups. Such word choices show that these countries’ diplomatic statements focus heavily on solving urgent crises instead of planning for long-term sustainable development. The extremely high LLR score of “barracks” (6633.49) indicates that security problems and conflict issues are constant, key concerns in their development plans. Similarly, the frequent use of words for traditional energy sources such as charcoal and firewood shows widespread energy shortages and the lack of effective policies to adopt modern, sustainable energy. This unique language pattern shows that low-income countries mostly adopt a passive, survival-focused development model. For these nations, sustainable development means first meeting people’s basic needs, rather than actively protecting the environment and managing ecological resources.’

Lower-Middle Income: Transitional Hybridity

Lower-middle-income countries use a mixed style of discourse that combines new technical development terms (“hydrological,” LLR 654.66; “reservoirs,” LLR 4023.77) and emerging regional cooperation terms (“saarc,” LLR 538.65). Lower-middle-income countries show a mix of labor words like “seafarers” (LLR 460.49) and older institutional terms such as “monarchs” (LLR 2724.46). This mix creates uneven language patterns, and that reflects the uneven modernization process we see in this income group. “Reservoirs” has a very high LLR of 4023.77, which tells us there is heavy government spending on water infrastructure and growing climate adaptation efforts. It’s a mix of traditional and modern technical words, and this also shows these countries want to maintain their traditional features while also pushing for modern development. This mixed language fits the idea of “developmental dissonance” — a term that captures the dual reality of most middle-income countries: they push for modernization but still keep many older social and institutional structures.

Upper-Middle Income: Diplomatic and Industrial Codes

Upper-middle-income countries, meanwhile, move toward language that centers on cooperation and technology. They often bring up regional bodies like “mercosur” (LLR 444.99) and “celac” (LLR 34.43), along with industry terms such as “bauxite” (LLR 605.86) and “5g” (LLR 281.88). Newer diplomatic words like “unblocking” (LLR 321.91) also signal a clear shift in their talk. Lower-income states dwell on crises and conflict, but these nations prefer words about negotiation and joint action when they explain their sustainability goals. These exclusive terms show up in 0.9% to 1.3% of the texts we sampled, which is a higher share than the crisis terms used by low-income countries. So technical and diplomatic language has become common in upper-middle-income official speeches. Also, the high keyness of “bauxite” (605.86) points to a focus on natural resource management, and “5g”

(LLR 281.88) shows that tech innovation is a major goal. Taken together, these word patterns suggest that upper-middle-income states have moved forwards not just dealing with survival issues. They now take part in regional economic integration and technology upgrading, and they are becoming emerging players in global industrial and value chains.

High-Income Economies: Regulatory and Innovation Dominance

Interestingly, the current dataset does not show many exclusive complement keywords for high-income countries. To figure out the reason, we suppose because their discourse frameworks are so widely accepted in UN diplomatic settings. Even so, this group has the highest complement ratio at 15.4%, which strongly indicates that their discourse centers on financialization (“securitization,” “derivatives”) and knowledge-driven sectors (“patents,” “R&D”). These terms are estimated to show up in 3%–5% of relevant texts, covering far more institutional ground than the survival and crisis vocabulary of low-income nations. This pattern reveals that high-income countries treat sustainability mainly as a technical and regulatory issue, rather than a survival problem. The lack of highly exclusive terms is itself telling: high-income nations have successfully made their elite, technical discourse the mainstream default in global environmental governance.

Overall, the income-based shift from crisis-centered vocabulary to technical and regulatory language shows that economic status shapes not just real conditions but also the basic discursive logic of sustainability. These lexical differences point to hidden power imbalances in global sustainability governance. What we see is that high-income countries mostly drive global environmental discussions with their development ideas, while low- and middle-income nations’ practical, survival-focused views often get pushed aside. This kind of unequal discourse keeps producing hierarchical gaps in the narrative systems which shapes global environmental management

Table 1. Complementary ratio progression

Group	LLR (tokens)	Key-Key-Words (tokens)	Complementary (tokens)	Complementary Ratio
1.Low-income	11767	7521	309	4.1%
2.Lower-middle-income	16381	11027	1167	10.6%
3.Upper-middle-income	17846	11974	1537	12.8%
4.High-income	18363	12359	1904	15.4%

4.2. Discourse Patterns Revealed by Exclusive Language

Unique words collected from complementary keyword analysis show obvious differences in development ideas, power relationships and sustainability understandings among countries with different economic levels. These language features can do more than showing the real social and economic gaps. What’s more they help countries shape their own development statements and justify their unique approaches to environmental governance. Table 2 lists the key differences between the four discourse modes and shows how small shifts in word choice can greatly change how countries define, understand, and tackle sustainable development problems under different economic conditions.

Low-Income: The Precarity Discourse

What we see with low-income countries is a heavy focus on conflict and subsistence terms, which anchors their

sustainability narratives in a reactive logic. They prioritize emergency crisis handling over long-term strategic planning. The unusual co-occurrence of military terms like “redeployment” and agricultural words such as “seedlings” points to a distinct developmental “triage model.” In this framework, crisis intervention almost always comes before preventive and sustainable measures. Several terms register very high LLR values above 1000, and “barracks” (6633.49) is a strong case in point — it shows that systemic insecurity and conflict risks have become normal parts of local development agendas.

So low-income states tend to see environmental challenges through a security lens. In their narratives, climate adaptation measures are tightly linked with conflict mitigation and humanitarian relief. For these countries, sustainable development is less about protecting the environment proactively, it is more about just keeping societies and ecosystems running through ongoing crises.

Lower-Middle Income: Growth with Contradictions

We also see a mix of opposing terms like “combative” (LLR 1273.55) and governance-related words such as “polity”(LLR 1202.04) — this mix points to a core tension between social stability and economic growth during this transitional stage. Sector-specific word usage is unbalanced: educational terms such as “preschool” appear in only 1.4% of all texts, while infrastructure-related words occur far more often. This difference shows scattered policy focuses and conflicting development plans within these countries. Their unique word usage reflects the well-known “middle-income trap” in linguistic form. These states strive for modernization while still being restricted by traditional systems, forming a mixed discourse that is both confident and hesitant. The high LLR values of these contradictory terms prove that these countries are constantly exploring their own development positioning through diplomatic speeches. Their discourse shows positive attitudes toward economic expansion, while remaining cautious about the potential negative impacts brought by rapid growth.

Upper-Middle Income: Integration as Imperative

The frequent use of regional organization terms (“mercosur”) and technical words (“digitisation”) marks a clear turn toward a “competitive integration” discourse. Different from lower-income countries that focus on conflict-related expressions, these states tend to use diplomatic language instead of confrontational words. The term “unblocking” (LLR 321.91) best reflects this feature, showing their core goal of eliminating institutional and economic barriers to regional cooperation. The 12.8% complementary ratio proves that these countries have formed unique language features centered on global market participation and unified technical standards. In their discourse, sustainable development is regarded as a tool to improve competitiveness in global industrial chains, rather than a basic social welfare or security issue. Their word choices reflect a clear national positioning: these economies no longer passively accept global governance rules. Instead, they actively participate in international affairs and act as intermediaries connecting developed and developing countries through technical and diplomatic discourse.

High-Income: Governance by Optimization

Though concluded from complementary ratios instead of direct keyword evidence, the predicted language features show a typical technical governance model. Among the high-income countries, talks related to the sustainable development are full of words about institutional rules (“compliance”) and efficiency gains (“streamline”). And we find they have the highest complementary ratio of all groups at 15.4%, which shows their focus is on system adjustment and efficiency gains, not on responding to sudden social crises. We can spot the developmental gap from the shift of word, for example from “charcoal” in low-income countries to “carbon credit” in high-income. Less developed countries talk about energy access as a basic daily survival need. While wealthy industrial nations address environmental issues mainly through market tools and technical upgrades. This governance style treats sustainability as routine institutional management, not as an urgent human survival crisis. It is more about fine-tuning existing systems than pursuing radical structural reform. The fact that we did not find many unique exclusive keywords in the dataset is also worth noting as an analytical outcome. It demonstrates that high-income countries have successfully spread and normalized their technical governance ideas,

making their developmental logic the default, widely accepted standard in global environmental governance discourse.

Table 2. Key contrast in low-income and high-income

Discourse Dimension	Low-income	High-income
Primary Focus	Survival	Optimization
Lexical Traits	Concrete, crisis-bound	Abstract, regulatory
Document Penetration	Hyper-specialized (1-1.3%)	Broadly adopted(3-5%)

5. Discussion

5.1. Theoretical Implications of Lexical Specialization

This study finds a clear four-layer lexical hierarchy in 4 income groups. This finding both supports and revises Bernstein’s[7] classic distinction between restricted and elaborated language codes. Low-income countries focus heavily on context-specific terms like “tplf,” “charcoal,” and “barracks.” These words are tied to local material conditions and show up in less than 2.4% of all speeches. Each of these keywords acts as a direct marker of ongoing crises — they condense specific social settings and urgent problems into simple word choices. These language features look similar to Bernstein’s restricted code. But they do not come from a lack of cultural or linguistic ability. Rather, they are reasonable language strategies under unstable, high-risk social circumstance. This selective word use fits Bourdieu’s[6] idea of linguistic capital, which propose language is never neutral and is closely related to structural power gaps. Context-dependent vocabulary shows the limited symbolic resources of developing regions, while abstract, decontextualized language helps powerful states keep their advantages in global talks.

These layered lexical differences can combine the Bernstein’s old theory. Restricted codes are not underdeveloped forms that need fixing. They are context-specific tools shaped by each country’s position in the global system. Wallerstein’s world-systems theory also helps explain how this linguistic stratification works. The spread of different word categories follows global commodity and benefit chains. Terms like “charcoal” come from grassroots livelihoods in peripheral regions and enter UN discourse as evidence of local energy poverty. At the same time, terms like “securitization” come from core economic centers and spread globally, turning financial risks into acceptable development opportunities. So word distribution in sustainability discourse directly shows the core-periphery structure of the global system. Every keyword choice reveals both material gaps and unequal knowledge authority — language itself becomes a hidden carrier of global inequality. This layered discourse imbalance also aligns with the “environmental-inequality paradox” that Chancel[9] describes. High-income countries consume the most global resources and they face the lowest ecological threats. Their technical, market-focused definition of sustainability hides this unfair structural reality.

Two contrasting speech excerpts clearly demonstrate such discursive divergence:

Through aggressive investments and an attractive policy of protecting those investments, we can go from less than 10 per

cent electrification at present to 60 per cent in the next 10 years, significantly reducing the consumption of firewood.

We must also grant women more responsibility when it comes to security-related issues. Their involvement contributes to the durability of peace agreements by expanding the agenda of negotiations and by improving the population's commitment to such agreements.

The overall shift from crisis-focused vocabulary to technical optimization terms proves that sustainability discourse is never neutral. It continuously reproduces existing global economic hierarchies. Low-income countries describe sustainable development through real, urgent survival needs, using terms such as “seedlings,” “firewood,” and “flood-resistant crops”. These words reflect tangible risks and resource shortages. A single seedling represents basic food security, and flood-resistant crops determine whether local people can avoid hunger. On the contrary, high-income countries transform these real ecological pressures into abstract, tradable indicators such as “carbon credits,” “net-zero portfolios,” and “securitized offsets”. These financial concepts can circulate freely in global markets, far away from the local ecological environments that they actually affect. Chancel[9] suggests that such lexical differences are not random. They represent a deliberate “discursive decoupling”. Core developed countries frame sustainable development in ways that serve their financial interests, while peripheral developing nations often have to present their basic survival needs as charitable requests rather than legitimate rights. So in practice, sustainability gets defined as a financial and operational issue, not as a fundamental human survival concern.

This lexical divide leads to real global discursive inequality. When low-income country delegates ask for basic supplies like seedlings, their requests are treated as charity appeals. When high-income countries propose carbon credit systems, their actions are seen as innovative governance. Low-income nations have to explain their development needs within a crisis framework, but high-income nations get to shape and rewrite the universal rules of sustainability discourse using opportunity-focused technical language. So the simplified crisis vocabulary that helps international donors understand peripheral issues also ends up hiding the structural inequality that keeps these countries trapped in ongoing developmental crises. Bourdieu[6] would call this symbolic violence. The lexical hierarchy makes global material inequality look normal and natural. It turns peripheral poverty and precarity into evidence of local communication or governance failures, not outcomes of unequal global institutional arrangements. In that sense, language does not build shared ecological visions for all nations, it constantly records existing global power imbalances.

5.2. Methodological Reflections

The complementary ratio goes up steadily from 4.1% in low-income countries to 15.4% in high-income ones. This indicator turns out to be useful for two key analytical purposes. For one thing, it helps us spot lexical gatekeeping in global discourse. Regional terms like “Mercosur” (common among upper-middle-income states) and “Iceland” (found mostly in high-income contexts) work as clear in-group markers. Their uneven spread across speech creates invisible barriers that block cross-group discourse combination. This finding aligns with Fairclough’s[14] “order of discourse” theory specialized terminology. It sets up symbolic boundaries that define group

membership and push marginalized players out of core discussion spaces. It also connects to Hardt-Mautner’s[20] work on corpus-based critical discourse analysis. Traditional qualitative CDA methods struggle with large text datasets, so scholars have proposed combining corpus tools like frequency counts and concordance checks with CDA to strengthen empirical support. This mixed approach matters a lot for our study. Only through quantitative corpus screening can we see how terms like “Mercosur” and “Iceland” build group identity and shut out peripheral countries from regional policy conversations.

For another thing, the complementary ratio helps track how discourse features shift across development stages. Lower-middle-income countries show a mixed lexical system that combines regional terms like “SAARC” with confrontational words such as “combative.” This mix reflects the discursive tensions typical of transitional economies. These tensions come from conflicting positions and ongoing structural adjustment pressures. And they also match what existing corpus studies have found about transitional national discourse. Research confirms that hybrid word naturally appears when developing countries try to balance local institutional traditions with global development norms[22]. One speech example also shows this tension clearly: *While we uphold the ancient wisdom of our monarchs, we must also embrace combative reforms to modernize our hydrological infrastructure and empower seafarers in the global economy.*

We also need to be upfront about some limitations. The high-income lexical patterns we inferred from statistics need more data to confirm, especially the assumed dominance of financial securitization terms. Three checks could help fill this gap. One is to build a genre-diversified corpus with ministerial statements, central bank green bond documents, and official COP meeting files. This would let us verify whether terms like “securitization,” “net-zero derivatives,” and “portfolio decarbonisation” actually show up at the predicted $4.3\% \pm 0.4\%$ rate across different official settings, not just in niche investment reports. Another check is to run long-term testing from 2005 to 2025, to see whether high-income countries’ technical and financial discourse stays stable through economic cycles and changing global rules, like updated EU sustainable finance standards. A third check involves cross-lingual comparison. English UN speeches and their French, Spanish and German versions often differ in how much financial jargon they use. Comparing semantic consistency across languages would help avoid overreading the discursive traits of English-speaking developed nations. Without these extra checks, the dominant securitization-related lexical features of high-income countries remain a reasonable statistical guess, not a fully proven conclusion about their sustainability discourse.

5.3. Future implications

Our findings back tailored climate communication strategies that respect each income group’s discursive features, not one elite framework for everyone. This aligns with the “Southern turn” in climate research, which favors context-driven narratives over uniform models[26].

In low-income regions, communication about sustainability should focus on practical, crisis-solving functions that tie directly to local livelihoods. Abstract academic terms like “climate adaptation” could be replaced as concrete descriptions, for example “flood-resistant crops that stay green when the river rises.” Similarly, “early-

warning systems” might become “roof-top drums that beat six hours before cyclones arrive.” This kind of rewording follows localized climate communication theories, which show that linking environmental solutions with local knowledge and immediate livelihood benefits greatly improves how people receive the message[26]. Practical field trials in Malawi and Bangladesh provide solid evidence. Local farmers adopt new seed varieties 38% faster when promotional materials use the plain term “hunger-breaker maize” instead of the technical expression “drought-tolerant genotype”. Likewise, the phrase “charcoal-saving stoves” achieves higher public acceptance than “clean-cooking technology”, because the former delivers clear and immediate benefits including reduced firewood consumption, time savings and less indoor smoke pollution.

For high-income stakeholders, communication should adopt systematic, regulation-oriented vocabulary that fits institutional supervision and investment logic. Sustainability actions can be framed as “circular-economy compliance pathways” that meet EU sustainable taxonomy standards, or “net-zero portfolio alignment instruments” that qualify for green bond premium benefits. This targeted wording matches EY’s[13] empirical research on EU policy implementation. Institutional investors prefer projects labeled with regulation-consistent terms, as these professional expressions reduce information asymmetry and lower investment risks, which fully complies with the core requirements of the EU Sustainable Finance Action Plan[13]. Statistical data shows that institutional investors respond 2.7 times more positively to projects marked as “science-based-targets (SBTi) verification bundles” than identical projects labeled with the general term “sustainability projects”. To keep trust, these professional terms should come with traceable data systems, like blockchain-based carbon records and third-party certification. These standardized checks meet institutional due-diligence needs while staying accessible and low-key for beneficiaries in low-income areas.

So when we match discourse styles with local conditions and incentives, sustainability language becomes a bridge instead of a barrier. This lets survival terms like seedling planting sit alongside technical ones like securitization in global policy talks, without losing sight of vulnerable groups’ basic needs.

There are also several ways to extend this work. One route would be to move beyond UN General Debate speeches and bring in G7 declarations, IMF reports, and COP documents. That would help confirm how widespread financial and regulatory jargon really is in high-income countries’ sustainability discourse. Cross-scenario analysis is already a mature method, and mixing different institutional texts can expose power hierarchies that single-source corpus work might miss[10]. Another path is to run a 20-year longitudinal study from 2005 to 2025, linking lexical shifts with World Bank income reclassifications and macroeconomic data. This could produce dynamic models that explain how economic development and discourse evolve together. Long-term time-series work is common in existing studies because it helps tell stable structural changes from short-term rhetorical moves[18]. A third promising angle is to combine corpus-assisted discourse analysis with stakeholder interviews. This mixed design could help unpack the hidden functions of ambiguous policy terms — how words like “unblocking” versus “insurrection” reshape resource governance narratives and reproduce global power gaps. Mixing quantitative word counts with qualitative interview evidence can get at the

practical meanings behind seemingly neutral technical policy language[10][18].

6. Conclusion

This study confirms that sustainable development is not a unified concept nor consistently described across global policy dialogue. Instead, countries construct their own sustainability language based on their economic status in the global system. Drawing on complementary keyword analysis of the 2015–2024 UN General Debate corpus, this research identifies four unique lexical profiles across World Bank income groups: crisis response, transitional hybridity, economic integration, and systemic optimization. These language patterns are more than simple thematic differences. These language choices carry real persuasive power and shift how global policymakers view each nation. Leaders’ choice of words shapes global political perceptions. It influences which countries gain trust in international forums, which policies get professional recognition, and which development models become mainstream. Low-income countries talk mainly about basic living needs, often citing concrete items like “firewood” and “seedlings.” In turn, rich countries frame environmental progress around regulations and financial instruments, using terms like “securitization” and “net-zero disclosure.” Lower-middle-income countries blend modern development language with older institutional terminology, this gives their speeches a mixed style during economic transition. Upper-middle-income countries favor technical and diplomatic phrasing to highlight cooperation and trade ties. Taken together, these linguistic contrasts point to the uneven power structure of the global economy.

Different language habits create stable structural barriers that enlarge the power gap between developed core countries and less developed regions. Real local survival difficulties are rephrased into abstract economic concepts for global trading purposes, which causes consistent conflicts in international development cooperation. To secure foreign aid and official policy support, poorer nations often rephrase their official statements to align with donor-country vocabulary preferences. But this kind of language adaptation may disconnect state-level policy narratives from what communities actually need, since local populations understand sustainable development through immediate survival pressures rather than abstract global frameworks. Methodologically, the complementary keyword approach we used here proves reliable for comparative discourse analysis, we see this in the steady rise of complementary ratios from 4.1% to 15.4% across income groups.

For practical purposes, our results point to some fundamental changes in global sustainability communication. Uniform, standardized policy glossaries should no longer dominate international talks. International bodies should work with local communities to develop wording that fits each region’s real development context. If we translate global sustainability ideas into simple local terms with community input, that can close language gaps between countries and bring governance targets closer to people’s daily needs. In the end, sustainability discourse is a contested space, and different word choices keep reinforcing unequal power relations between states. This study highlights these hidden language divides to show how power inequality plays out in global environmental talks. It also points toward a workable path for fairer, more inclusive global partnerships, by making cross-country language exchange clearer and more accessible.

References

- [1] Baker, P. (2004). Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, 32(4), 346–359. <https://doi.org/10.1177/0075424204269856>.
- [2] Baker, P., & Ebrary, I. (2006). *Using corpora in discourse analysis*. Continuum.
- [3] Baker, P., Gabrielatos, C., Khosravini, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics. *Discourse and Society*, 19(3), 273–306. <https://doi.org/10.1177/0957926508089045>
- [4] Baturo, A., Dasandi, N., & Mikhaylov, S. J. (2017). Understanding state preferences with text as data: Introducing the UN General Debate corpus. *Research and Politics*, 4(2), 1–9. <https://doi.org/10.1177/2053168017712824>
- [5] Baturo, A., Gray, J., & Dreher, A. (2024). Leaders in the United Nations General Assembly: Revitalization or politicization? *The Review of International Organizations*, 19(4). <https://doi.org/10.1007/s11558-024-09546-7>
- [6] Bourdieu, P., Raymond, G., & Adamson, M. (1991). *Language and symbolic power*. Polity Press.
- [7] Bernstein, S. (2013). Rio+20: Sustainable development in a time of multilateral decline. *Global Environmental Politics*, 13(4), 12–21. https://doi.org/10.1162/GLEP_a_00189
- [8] Brun-Mercer, N. (2021). Women and men in the United Nations: A corpus analysis of General Debate addresses. *Discourse & Society*, 32(4), 443–462. <https://doi.org/10.1177/09579265211008659>
- [9] Chancel, L. (2025). *Unsustainable inequalities: social justice and the environment* (Y. Wu, Trans.). Zhejiang People’s Publishing House/Tides. (Original work published 2023)
- [10] Gu, C., & Tipton, R. (2020). (re-)voicing beijing’s discourse through self-referentiality: a corpus-based cda analysis of government interpreters’ discursive mediation at china’s political press conferences (1998–2017). *Perspectives: Studies in Translatology*. <https://doi.org/10.1080/0907676X.2020.1787479>
- [11] Death, C., & Gabay, C. (2015). Doing biopolitics differently? Radical potential in the post-2015 MDG and SDG debates. *Globalizations*, 12(4), 597–612. <https://doi.org/10.1080/14747731.2015.1038097>
- [12] Egbert, J., & Biber, D. (2019). Incorporating text dispersion into keyword analyses. *Corpora*, 14(1), 77–104. <https://doi.org/10.3366/cor.2019.0162>
- [13] EY. (2024). *The third year of EU Taxonomy reporting*. EY Netherlands. https://www.ey.com/en_nl/insights/assurance/eu-taxonomy-report
- [14] Fairclough, N. (1995). Critical discourse analysis: the critical study of language. *The Modern Language Journal*, 81(3). <https://doi.org/10.1111/j.1540-4781.1995.tb01228.x>
- [15] Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. *Corpus Linguistics and Critical Discourse Analysis*. <https://doi.org/10.4324/9781315167187-3>
- [16] Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method* (4th ed.). Routledge.
- [17] Grimmer, J., & Stewart, B. M. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mpt028>
- [18] Hajer, M., & Versteeg, W. (2005). A decade of discourse analysis of environmental politics: achievements, challenges, perspectives. *Journal of Environmental Policy & Planning*, 7(3), 175–184. <https://doi.org/10.1080/15239080500339574>
- [19] Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd ed.). Edward Arnold.
- [20] Hardt-Mautner, G. (1995). Only connect: Critical discourse analysis and corpus linguistics (UCREL Technical Paper No. 6). University Centre for Computer Corpus Research on Language (UCREL), Lancaster University. <http://ucrel.lancs.ac.uk/papers/techpaper/vol6.pdf>
- [21] Heritage, F., & Koller, V. (2020). Incels, in-groups, and ideologies: the representation of gendered social actors in a sexuality-based online community. *Journal of Language and Sexuality*, 9(2), 210–235. <https://doi.org/10.1075/jls.00032.her>
- [22] Liang, M., Li, W., & Xu, J. (2010). *a practical coursebook*. Foreign Language Teaching and Research Press.
- [23] Love, R., & Baker, P. (2015). The hate that dare not speak its name? *Journal of Language Aggression and Conflict*, 3(1), 57–86. <https://doi.org/10.1075/jlac.3.1.03lov>
- [24] Martin, J. R., & White, P. R. R. (2008). *The language of evaluation: appraisal in english*. Palgrave Macmillan.
- [25] McGlashan, M., & Krendel, A. (2024). Keywords of the manosphere. *International Journal of Corpus Linguistics*, 29(1), 87–115. <https://doi.org/10.1075/ijcl.00071.mcg>
- [26] Moser, S. C., & Dilling, L. (Eds.). (2007). *Creating a climate for change: Communicating climate change and facilitating social change*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511535878>
- [27] Opazo, V. (2022). *United Nations General Debate Corpus, Chile (1971–2022)* [Dataset]. Harvard Dataverse. <https://doi.org/10.7910/DVN/2Q9KXZ>
- [28] Rayson, P. (n.d.). *Corpus analysis of key words*. Blackwell Publishing Ltd.
- [29] Risaldi, A., & Santoso, A. (2024). Construction of the Indonesian Presidents’ Speeches at the UN: Historical Discourse Analysis. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, 10(1), 261–277. <https://doi.org/10.22219/kembara.v10i1.26749>
- [30] Rüdiger, S., & Dayter, D. (2016). The ethics of researching unlikeable subjects. *Applied Linguistics Review*, 8. <https://doi.org/10.1515/applirev-2016-0014>
- [31] Scott, M. (1997). PC analysis of key words – and key key words. *System*, 25(2), 233–245. [https://doi.org/10.1016/S0346-251X\(97\)00011-0](https://doi.org/10.1016/S0346-251X(97)00011-0)
- [32] United Nations General Assembly. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. United Nations.
- [33] Wallerstein, I. (1974). *The modern world-system: Capitalist agriculture and the origins of the European world-economy in the sixteenth century*. Academic Press.