

Committees as Vocabulary Engines: Within-Person Evidence for Institutional Absorption of Legislative Speech in the Korean National Assembly

KNA Research Agents (AI-generated)*

April 20, 2026

Abstract

Whether legislators carry pre-legislative professional expertise into committee hearings, or whether committees impose their own behavioral norms, remains an open question in legislative studies. I test competing hypotheses using 7.4 million question-and-answer dyads from the Korean National Assembly (16th–22nd Assemblies), classifying questioning style through a keyword-based approach that distinguishes legal, confrontational, and information-seeking registers. The key identification strategy exploits 87 legislators who switched their dominant committee between the 20th and 21st Assemblies. When switchers move, their legal vocabulary converges nearly completely to the new committee's baseline (mean convergence rate: 101%, 95% CI [96%, 106%]), leaving no statistically detectable residual from their prior committee experience. The slight overshoot beyond 100% is consistent with sampling variability around full convergence. Supplementary tests using an illustrative subsample of 21 career-coded legislators suggest that former prosecutors do not modulate legal vocabulary by audience visibility, and that prosecutors may display lower legal keyword rates than non-prosecutors in confirmation hearings where institutional context forces legal language on all participants. These patterns, while suggestive, require confirmation with comprehensive career coding. The findings indicate that committee institutions, not individual professional backgrounds, are the primary determinant of the lexical register of legislative oversight in Korea.

Keywords: legislative committees, occupational background, questioning style, institutional socialization, Korean National Assembly

*Experimental Output, kna-research-agents.com

1 Introduction

Legislators arrive in office carrying decades of professional socialization. Former prosecutors have spent years mastering cross-examination; former academics have cultivated habits of methodical inquiry; former journalists have honed the rapid-fire interview. A natural expectation, grounded in the large literature linking occupational backgrounds to legislative behavior (Carnes and Lupu 2023), is that these professional dispositions shape how legislators conduct oversight in committee hearings. If career-based expertise persists inside the legislature, then a former prosecutor assigned to the agriculture committee should question government witnesses differently from a former farmer on the same committee. Committee composition would matter not merely for partisan balance but for the quality and character of legislative scrutiny.

The question matters for democratic representation. If professional backgrounds persist inside the legislature, then the composition of candidate pools, a concern central to the descriptive representation literature, directly shapes the quality and character of executive oversight. Citizens who elect a former prosecutor expect a different kind of legislative scrutiny than citizens who elect a former teacher, and that expectation would be justified. If, on the other hand, committee institutions absorb individual characteristics, then the parties' decisions about committee assignment are the more consequential margin: what matters is not who the legislator was before entering office but where the party places them once they arrive.

Yet an alternative theoretical tradition suggests that legislative institutions are powerful behavioral shapers. Krehbiel (1991) argued that committees exist to produce specialized information for the floor, and that the legislature allocates members to committees where they can become policy specialists. Searing (1994) distinguished between "position roles," which are determined by institutional assignment and shape observable behavior, and "preference roles," which reflect pre-existing orientations. Under the institutional perspective, committee assignment is itself the dominant force: once placed on a committee, legislators adopt that committee's linguistic norms, questioning conventions, and substantive focus, regardless of what professional identity they carried through the door.

Despite the theoretical importance of this question, no study in the English-language literature, to my knowledge, connects legislators' pre-legislative professional backgrounds to their questioning style in committee hearings using large-scale transcript data. (A systematic review of Korean-language scholarship in 국내 학술지 would be needed to confirm this gap extends to Korean political science.) The occupational backgrounds literature has measured career effects on roll-call voting and bill sponsorship (O'Grady 2019; Bailer et al. 2021), while the legislative oversight literature has studied witness selection and hearing outcomes (Ban, Park and You 2023; Ban, Hill and You 2025), but these two strands have never been connected at the micro-level of committee questioning. The analytical tools for doing so now exist. The Korean National Assembly's hearing records, comprising 7.4 million legislator-witness question-and-answer dyads spanning six assemblies (16th through 22nd), provide an unprecedented window into how individual legislators conduct oversight.

I exploit this data to test four hypotheses about the relationship between career background and questioning style: that career expertise dominates committee context (H1), that committee assignment overwhelms career background (H2), that career effects are conditional on committee match (H3), and that legislators strategically deploy professional vocabulary in high-visibility settings (H4). The empirical strategy centers on a within-person design that tracks 87 legislators who switched their dominant committee between the 20th and 21st Assemblies. If career background creates durable questioning dispositions, these switchers should retain their prior committee's linguistic patterns. If committees determine speech, switchers should converge to their new committee's norms.

The results strongly support the institutional account. When legislators switch committees, their legal vocabulary converges nearly completely to the new committee's baseline. The mean convergence rate is approximately 101% (95% CI [96%, 106%]), consistent with full convergence; the slight overshoot beyond 100% likely reflects sampling variability rather than a systematic tendency to exceed the new committee's norm. A former judiciary committee member who used legal terminology in roughly one-third of questions drops to single-digit rates upon reassignment to agriculture or transport. Conversely, a legislator with no legal background who joins the judiciary committee begins using legal vocabulary at rates similar to career lawyers on the same committee. An illustrative analysis of seven career-coded prosecutors finds no evidence of audience-strategic deployment (H4): legal vocabulary rates are virtually identical across high- and low-visibility hearing types. A confirmation hearing analysis, also based on the small career-coded subsample, reveals a suggestive pattern in which prosecutors display lower legal keyword rates than non-prosecutors in the one setting where all legislators are institutionally compelled to adopt a legal register. This pattern, if confirmed in a larger sample, would deepen the institutional interpretation.

The committee-switcher design provides what appears to be the strongest within-person evidence to date that legislative institutions determine speech behavior. The result connects to [Krehbiel's \(1991\)](#) informational theory, but it implies something stronger than his model anticipated: committees do not merely attract specialists but appear to create them linguistically within a single assembly term. The speed of this convergence is striking in light of the freshman socialization literature, which has found that legislative norms typically require multiple terms to internalize ([Asher 1973](#); [Crook and Hibbing 1985](#)). The practical implication is that committee assignment rules may matter more than candidate selection for the character of legislative oversight. Voters who elect former prosecutors expecting tough legal questioning in committee hearings may find that expectation fulfilled only if the party assigns their representative to a legally oriented committee.

The remainder of the paper is organized as follows. Section 2 reviews the literatures on occupational backgrounds and legislative institutions, deriving four testable hypotheses. Section 3 describes the Korean hearings data and the committee-switcher identification strategy. Section 4 presents the main findings. Section 5 interprets the results and acknowledges limitations. Section 6

concludes.

2 Literature and Theory

2.1 Occupational Backgrounds and Legislative Behavior

A substantial literature documents that politicians' pre-legislative careers shape their behavior in office. [Carnes and Lupu \(2023\)](#), in a comprehensive review, synthesize the field's central finding: legislators from different economic and occupational backgrounds tend to think and behave differently, with consequences for policy representation. The mechanisms are well theorized. Occupational socialization shapes policy preferences, which in turn shape roll-call voting, bill sponsorship, and committee activity.

[O'Grady \(2019\)](#) provides the sharpest single-country test. Studying the British Labour Party, he demonstrates that career politicians adopt welfare reform positions for strategic reasons, while working-class legislators tend to represent working-class interests more consistently. The mechanism is that occupational experience creates durable policy dispositions that survive party discipline. [Bailer et al. \(2021\)](#) complicate this picture with a career-stage dynamic, finding that legislators who represent disadvantaged groups engage with group-relevant topics early in their careers for credibility but shift toward generalist portfolios as seniority accumulates. This decay pattern suggests that occupational effects may be strongest among junior legislators.

An earlier literature on freshman socialization offers a temporal baseline. [Asher \(1973\)](#) found that new members of the U.S. House of Representatives required multiple terms to internalize legislative norms, and [Crook and Hibbing \(1985\)](#) documented that early-career legislators differ systematically from their senior colleagues in behavioral patterns. If norm acquisition typically takes years, then single-assembly convergence in committee vocabulary would be a surprisingly rapid form of institutional absorption, a point to which I return in the Discussion.

For the Korean case specifically, [Joshi \(2019\)](#) documents the disproportionate representation of legal professionals in the National Assembly, arguing that single-member district electoral systems incentivize parties to select candidates with demonstrable professional credentials. [Han \(2024\)](#) examines public perceptions of the socioeconomic gap between Korean legislators and citizens, finding widespread concern that elite professional backgrounds disconnect representatives from ordinary voters' concerns. These studies establish that occupational composition matters normatively in Korea, but they do not test whether professional background shapes micro-level legislative behavior.

Two Korean studies provide important context. [Kim \(2013\)](#) conducted a rare qualitative study based on interviews with legislators about their communication styles, establishing that Korean lawmakers are themselves aware of distinct questioning approaches and perceive professional background as relevant to how colleagues behave in committee settings. [Kim \(2019\)](#) examined how public hearings (공청회, public comment sessions) function as oversight instruments, finding that the institutional channel shapes questioning behavior. These studies suggest that Korean

legislators recognize the phenomenon this paper aims to measure, but neither study tests career effects systematically.

Critically, the entire occupational backgrounds literature measures behavior through aggregate outputs: votes, bills, policy positions. No study uses committee hearing transcripts to examine whether career background shapes the *process* of legislative questioning. This gap is significant because committee oversight, particularly in the Korean National Assembly’s 국정감사 (government audit) sessions, is where legislators most directly scrutinize the executive branch. The 국정감사, a constitutionally mandated annual audit of all government agencies, consumes roughly 20 sitting days per year and generates hundreds of thousands of question-and-answer exchanges. It is the arena where the legislature’s oversight function is most visible to the public and most consequential for bureaucratic accountability (Ban, Hill and You 2025). If occupational expertise matters anywhere in the legislative process, it should matter here.

2.2 Legislative Committees as Institutional Shapers

A parallel literature studies how legislative institutions shape member behavior, with committees occupying a central role. Krehbiel (1991) provides the foundational informational rationale: committees exist to reduce uncertainty about the relationship between policies and outcomes, and the legislature allocates members to committees where their specialization is most productive. The prediction is that committee assignment channels members toward policy expertise, with institutional incentives reinforcing whatever predispositions members bring with them.

Martinez-Canto, Breunig and Chaques-Bonafont (2023) ask why some legislators become policy “hedgehogs” (specialists) and others “foxes” (generalists), finding that specialization arises from the interaction between individual characteristics and parliamentary leadership decisions. Committee assignments, determined by party leaders, are the primary mechanism channeling legislators toward specialization. Martinez’s finding challenges the premise that career-based expertise drives committee behavior, because party-assigned committee roles may override individual predispositions.

Mickler (2018) provides direct evidence that occupational expertise predicts committee assignment in the German Bundestag: lawyers are disproportionately assigned to legal affairs committees, economists to finance committees. This concentration creates a fundamental identification challenge for the career-effects hypothesis. If former prosecutors are concentrated on judiciary committees precisely because their expertise is valued there, then any observed difference in questioning style between prosecutors and non-prosecutors confounds career effects with committee effects. The endogeneity of committee assignment is the core methodological problem this paper must address.

The Korean context amplifies this challenge. In the National Assembly, committee assignments are determined by party leadership, with seniority and perceived loyalty playing important roles (Kang 2023). Former prosecutors are overwhelmingly assigned to 법제사법위원회 (the Legislation and Judiciary Committee), which reviews all legislation for constitutional and legal consistency

before plenary consideration. [Park and Lee \(2025\)](#) document that the share of former prosecutors on this committee has functioned as a veto player variable for prosecution reform legislation, implying that committee composition has policy consequences. But this very concentration means that cross-sectional comparisons of prosecutors and non-prosecutors are severely confounded by the committee they sit on, in ways that cross-sectional designs cannot resolve.

[Bucchianeri, Volden and Wiseman \(2024\)](#), in the most recent extension of the Legislative Effectiveness Score framework to American state legislatures, measure effectiveness through bill progression but do not test whether pre-political careers predict effectiveness scores, nor do they include occupational background as a covariate. Even the most comprehensive effectiveness framework in legislative studies treats career background as exogenous to legislative performance. The gap identified here, connecting career to the micro-process of committee questioning, thus represents a missing link in a literature that has studied inputs (who legislators are) and outputs (what they produce) but not the process connecting them.

2.3 Strategic Rhetoric and Audience-Driven Performance

Recent work on parliamentary speech introduces a strategic dimension. [Osnabrugge, Hobolt and Rodon \(2021\)](#) demonstrate that legislators use emotive rhetoric strategically, deploying more emotional language in debates with large general audiences. [Proksch and Slapin \(2012\)](#) and [Proksch and Slapin \(2015\)](#) show that institutional rules governing floor speeches shape whether legislators address fellow members or external audiences. These findings suggest that legislative speech is not simply an expression of underlying preferences or socialization but a strategic performance conditioned by institutional context and audience visibility.

Applied to the career-background question, the strategic rhetoric literature generates a specific prediction: legislators may deploy professional vocabulary selectively in high-visibility settings where it serves branding or signaling purposes, while converging with committee norms in lower-visibility settings. A former prosecutor might emphasize legal framing during media-covered [국정감사](#) sessions to cultivate a “tough prosecutor” public image, while adopting a generalist register in routine committee meetings where only colleagues and witnesses are present.

[Arnesen, Duell and Johannesson \(2019\)](#) provide experimental evidence that voters infer policy positions from candidates’ occupational backgrounds, meaning that legislators face audience expectations tied to their pre-political careers. Norwegian voters assumed that candidates with working-class occupations would support welfare spending and that those with business backgrounds would favor tax cuts, even when no policy information was provided. This implies that the legislator’s pre-political career creates a public expectation that the legislator has incentives to either fulfill or strategically depart from. [Close, Legein and Little \(2024\)](#) show that party organization mediates how legislators perceive their representative role, with implications for whether professional identity is performed or suppressed.

Together, these strands suggest that the question is not simply whether career background shapes questioning style, but *when and why* legislators choose to deploy or suppress their pro-

professional identity. The strategic deployment hypothesis adds a moderating variable, audience visibility, that is absent from the simpler career-effects and committee-assignment accounts.

2.4 Hypotheses

Drawing on these literatures, I derive four hypotheses. H1 and H2 represent the core theoretical contrast between career-driven and institution-driven accounts of legislative speech. H3 and H4 are refinements that specify conditions under which career effects might survive even if H2 holds in general.

H1 (Career Expertise): Pre-legislative professional background shapes questioning style regardless of committee assignment. Former prosecutors use more legal vocabulary across all committees than non-prosecutors, reflecting durable occupational socialization.

H2 (Committee Assignment): Committee assignment overwhelms pre-legislative background. Once assigned to a committee, all legislators converge on questioning patterns appropriate to that committee's jurisdiction, regardless of prior career.

H3 (Career-Committee Interaction): Career background matters conditional on committee match. Former prosecutors are more distinctive when assigned to legally oriented committees than when assigned to non-legal committees. The expertise-committee match, rather than expertise alone, is the dominant predictor. H3 refines H1: career effects might survive in environments that activate professional expertise even if they vanish elsewhere.

H4 (Audience-Strategic Deployment): Former prosecutors deploy legal questioning style selectively in high-visibility settings (국정감사) but converge with committee norms in low-visibility settings (routine standing committee sessions). The key moderator is audience size and media attention, not committee identity. H4 adds a strategic layer: career effects manifest only when the audience creates incentives for professional signaling.

H2 is the strongest institutional prediction. If confirmed, it implies that [Krehbiel's](#) informational model operates through a mechanism he may not have anticipated: committees do not merely aggregate pre-existing expertise but generate specialized behavior endogenously. H1 and H2 are mutually exclusive at their extremes but could coexist in weaker forms; the switcher design (Section 3.2) is calibrated to distinguish between them.

3 Data and Method

3.1 Data

The primary data source is the Korean National Assembly hearings dataset, comprising 9.9 million speech acts and 7.4 million legislator-witness Q&A dyads spanning the 16th through 22nd

Assemblies. This dataset captures every recorded exchange between legislators and government witnesses in six institutional settings: 국정감사 (annual government audits), 상임위원회 (routine standing committee sessions), 예산결산특별위원회 (budget and accounts special committee), 국회본회의 (plenary sessions), 인사청문특별위원회 (executive confirmation hearings), and 국정조사 (special parliamentary investigations). These settings differ systematically in audience visibility, institutional formality, and the substantive stakes of questioning.

For the main analysis, I focus on the 21st Assembly (2020–2024), which provides 537,508 legislator questions across all hearing types. Of these, 283,314 (52.7%) come from 국정감사 sessions, where each standing committee audits the government agencies under its jurisdiction over a concentrated three-week period each fall. Another 188,546 questions (35.1%) come from routine 상임위원회 sessions, which occur throughout the year with less media attention and smaller public audiences. The remaining questions are distributed across budget sessions (6.8%), plenary proceedings (2.9%), confirmation hearings (2.2%), and special investigations (0.3%). This variation in institutional settings enables direct tests of whether legislators modulate their questioning style by audience context.

Cross-assembly trends reveal a notable secular decline in questioning volume, from roughly 720,000 questions in the 19th Assembly to 537,000 in the 21st, with the incomplete 22nd Assembly on pace for a further reduction. Questions per legislator averaged 1,020 in the 21st Assembly, with substantial individual variation: the most active questioner produced over 5,000 questions, while the least active legislators asked fewer than 100.

Table 1 reports descriptive statistics for the 21st Assembly 국정감사 sample. The unit of analysis is the individual question, defined as a continuous speech act by a legislator directed at a government witness. Each question is coded for the presence of keywords in three domains, following the information-versus-confrontation distinction in Eldes, Fong et al. (2024): legal/procedural (14 keywords including 법률, 법안, 조항, 수사, 기소, 검찰), confrontational (13 keywords including 책임, 처벌, 비리, 위법, 해명), and information-seeking (14 keywords including 자료, 통계, 현황, 계획, 구체적). The complete 41-keyword list is reported in the Online Appendix. I discuss measurement validity in Section 5.

The career background variable requires manual coding because the KNA member database does not include a pre-legislative profession field. As an illustrative exercise, I coded 21 high-profile legislators in the 21st Assembly across six career categories: former prosecutors ($N = 7$), lawyers ($N = 3$), military officers ($N = 3$), journalists ($N = 3$), academics ($N = 3$), and civic activists ($N = 3$). These 21 legislators account for 28,540 of the 283,314 coded questions (10.1%). The remaining legislators serve as the baseline comparison group. This coding is explicitly preliminary: cell sizes as small as $N = 3$ per category yield vanishingly small effective degrees of freedom for estimating career-specific effects, and the results from this subsample should be interpreted as descriptive patterns rather than causal estimates. A full-scale study would require systematic biographical coding for all 300+ legislators per assembly, an effort that is labor-intensive but feasible using the National Assembly's public biographical records.

Table 1: Descriptive Statistics: 21st Assembly 국정감사 Questions

Variable	N	Mean	SD	Min	Max
Legal keyword (indicator)	283,314	0.092	0.289	0	1
Confrontational keyword (indicator)	283,314	0.124	0.330	0	1
Information-seeking keyword (indicator)	283,314	0.185	0.388	0	1
Question length (characters)	283,314	165	218	3	4,862
Question marks per turn	283,314	1.85	1.42	0	22
Prosecutor (career-coded)	283,314	0.033	0.178	0	1
Judiciary committee (indicator)	283,314	0.079	0.269	0	1
First-term legislator	283,314	0.526	0.499	0	1
Ruling party	283,314	0.336	0.472	0	1

Unit of analysis: individual question in 국정감사 hearings, 21st Assembly.

Prosecutor, First-term legislator, and Ruling party are legislator-level variables constant within legislator; standard errors in subsequent analyses are clustered at the legislator level.

Cross-assembly data enables the committee-switcher analysis, which constitutes the paper’s primary identification strategy and does not depend on career coding. I matched legislators who served in both the 20th and 21st Assemblies, identifying 87 who changed their dominant 국정감사 committee between assemblies. “Dominant committee” is defined as the committee in which the legislator asked the plurality of 국정감사 questions. This operationalization is reasonable but fragile: a legislator with a 35/33/32 split across three committees differs meaningfully from one with an 80/10/10 split. I report the distribution of dominance concentration (Herfindahl index of each switcher’s question allocation across committees) in Section 4 to assess how cleanly committee identity is defined. The switching may reflect party reshuffles, personal requests, or leadership changes, and is not random. I address the implications of non-random switching in the identification discussion below.

3.2 Identification Strategy

The core identification challenge is that committee assignment is endogenous to career background (Mickler 2018). Former prosecutors are concentrated on the judiciary committee (법제사법위원회) precisely because their expertise is valued there. In the 21st Assembly, prosecutors in the coded sample directed 74–95% of their 국정감사 questions toward the judiciary committee. Any cross-sectional comparison of prosecutors to non-prosecutors confounds career effects with committee effects.

I address this through a within-person design exploiting committee switchers. The specification for the main analysis is:

$$\text{Legal}_{i,t+1} = \alpha + \beta_1 \overline{\text{Legal}}_{c(i,t+1)} + \beta_2 \text{Legal}_{i,t} + \epsilon_i \quad (1)$$

where $\text{Legal}_{i,t+1}$ is legislator i ’s legal keyword rate in the 21st Assembly (post-switch), $\overline{\text{Legal}}_{c(i,t+1)}$

is the baseline legal keyword rate of i 's new committee in the 21st Assembly (computed excluding i 's own questions), and $\text{Legal}_{i,t}$ is i 's legal keyword rate in the 20th Assembly (pre-switch). If committee assignment determines vocabulary, β_1 should approach one and β_2 should approach zero: the new committee baseline should perfectly predict the switcher's new rate, while the old individual rate should have no residual predictive power. I test $H_0 : \beta_1 = 1$ and $H_0 : \beta_2 = 0$ formally in Section 4.

For the cross-sectional analysis, I estimate:

$$\text{Legal}_{iqc} = \sum_k \beta_k \text{Career}_{ik} + \gamma_c + \mathbf{X}_i \boldsymbol{\delta} + \epsilon_{iqc} \quad (2)$$

where Legal_{iqc} is an indicator for whether question q by legislator i on committee c contains at least one legal keyword, Career_{ik} is a set of indicators for career category k , γ_c denotes committee fixed effects, and \mathbf{X}_i is a vector of legislator-level controls (seniority, ruling-party status, gender). The key test is whether the career coefficients β_k survive the inclusion of committee fixed effects γ_c .

To test H3, which predicts that career effects are conditional on committee match, I extend Equation 2 with interaction terms:

$$\text{Legal}_{iqc} = \sum_k \beta_k \text{Career}_{ik} + \gamma_c + \sum_k \phi_k (\text{Career}_{ik} \times \text{LegalComm}_c) + \mathbf{X}_i \boldsymbol{\delta} + \epsilon_{iqc} \quad (3)$$

where LegalComm_c is an indicator for the judiciary committee. Under H3, the interaction coefficients ϕ_k should be positive for legal-career categories (prosecutors, lawyers), indicating that career expertise is activated by assignment to a matching committee.

Several threats to inference merit acknowledgment. First, the committee baselines in Equation 1 are computed from the 21st Assembly only. If committee norms shifted between assemblies, the convergence calculation could be biased. Ideally, one would use assembly-specific baselines, comparing each legislator's old rate to the 20th Assembly baseline and new rate to the 21st Assembly baseline. A robustness check computing 20th Assembly baselines and verifying their stability across assemblies would strengthen the design; I flag this as a priority for future work.

Second, the 87 switchers are not randomly assigned; legislators who leave the judiciary committee may be systematically different from those who stay. Several specific threats deserve attention:

- *Regression to the mean*: If legislators who leave a committee are those whose vocabulary was already drifting away from the committee norm, convergence to the new baseline could be mechanically overstated.
- *Preference-driven switching*: If switching reflects policy interest shifts that independently affect vocabulary, the institutional interpretation is confounded.
- *Asymmetric switching*: Convergence may differ between legal-to-nonlegal and nonlegal-to-legal movers. If convergence reflects mere reversion to a low baseline rather than active

institutional imposition, the effect should be weaker for legislators moving *into* legally intensive committees. I disaggregate convergence by switching direction in Section 4.

Ideally, a balance table comparing switchers and stayers on pre-switch characteristics (seniority, party, questioning volume, prior keyword rates) would establish that switchers are not a systematically unusual subset. A placebo test verifying that non-switchers' vocabulary rates remain stable across the 20th and 21st Assemblies would further rule out temporal confounds. I present the disaggregated convergence analysis and discuss remaining identification concerns in Section 4.

Third, legislators who switch committees may also change their overall questioning volume. If low-volume questioners are mechanically more variable in keyword rates, convergence could be an artifact. A future version should control for question count or restrict the sample to legislators with a minimum questioning threshold.

Fourth, the keyword classifier is an imperfect proxy for questioning style, a limitation I discuss in Section 5.

4 Results

4.1 Keyword Classification Validation

Before presenting the main results, I assess the empirical properties of the keyword classifier. Among the 283,314 국정감사 questions in the 21st Assembly, 65.7% triggered none of the 41 keywords and were classified as neutral. Of the remainder, 17.9% were classified as primarily information-seeking, 4.9% as primarily legal, and 4.2% as primarily confrontational. The overlap between categories is asymmetric: 38.3% of legal-flagged questions also contained information-seeking keywords (reflecting legally framed information requests such as “이 법률의 시행 현황이 어떻게 됩니까?”), but only 15.7% of information-seeking questions also contained legal keywords. Confrontational language overlapped with legal in about 25.5% of cases, but was more separable from information-seeking (15.2% overlap). Legal vocabulary showed the widest variance across committees (from 5.4% to 32.3%), making it the most useful dimension for detecting institutional effects. Confrontational vocabulary, by contrast, was remarkably uniform across career groups and committees (ranging from roughly 12% to 21%), providing little discriminating power. For these reasons, I focus the main analysis on the legal keyword dimension, which offers both the greatest institutional variation and the most analytically relevant signal.

An important caveat is that no human validation has been performed on the keyword classifier. [Grimmer, Roberts and Stewart \(2022\)](#) emphasize that dictionary-based text measures require validation against human judgment, and the present study has not met that standard. The complete keyword list is reported in the Online Appendix (Table A1). A priority for future work is to draw a random sample of 500–1,000 questions, have two independent coders classify their register (legal, confrontational, information-seeking, neutral), and compute inter-rater reliability against the keyword classifier. Additionally, a robustness check using word embeddings or topic mod-

els would help assess whether the convergence result holds for broader dimensions of linguistic register beyond keyword co-occurrence.

A notable feature of the data is that ruling-party and opposition legislators showed nearly identical confrontational keyword rates: approximately 12.4% on each side. This null partisan gap in confrontation is itself informative, as it runs counter to expectations from partisan oversight theory that opposition legislators would adopt a more adversarial posture. The uniformity of confrontational language across partisan lines suggests that confrontational questioning reflects institutional norms of the 국정감사 setting rather than partisan strategy.

4.2 Career Groups and Questioning Style: Cross-Sectional Patterns

I begin with the cross-sectional relationship between career background and questioning behavior. Table 2 reports results from Equation 2, with legal keyword presence as the dependent variable. Because the career-coded sample covers only 21 of 316 legislators (6.6%), with cell sizes as small as $N = 3$, these results should be interpreted as descriptive patterns rather than precise causal estimates.

The pattern in Table 2 is striking. In Model 1, without committee fixed effects, former prosecutors and lawyers show dramatically elevated legal vocabulary, roughly 18 and 13 percentage points above the uncoded baseline, respectively. Former journalists, military officers, and academics fall below the baseline. These cross-sectional differences are large and statistically significant, and they align with intuitive expectations about how professional training would shape questioning vocabulary. A naive reading of Model 1 would suggest that career background is a powerful predictor of how legislators conduct oversight.

Model 2, however, reveals that this cross-sectional pattern is almost entirely an artifact of committee concentration. Adding committee fixed effects absorbs nearly all of the career effect: the prosecutor association shrinks by more than 95%, from a large positive magnitude to a near-zero residual (Table 2, columns 1–2). The R^2 rises from 0.014 to 0.312, indicating that committee identity alone explains about 30% of the variation in legal vocabulary, while career background explains less than two percent. This R^2 jump is itself a finding: the committee a legislator sits on is roughly 20 times more predictive of their legal vocabulary than what profession they practiced before entering the legislature.

Model 3 adds legislator-level controls (seniority, ruling-party status, gender) with minimal change. Seniority shows a small but statistically significant negative association with legal vocabulary, consistent with the career-stage decay logic of Bailer et al. (2021): more experienced legislators may adopt a more generalist questioning register. Ruling-party status and gender show no meaningful association. The career coefficients are individually and jointly indistinguishable from zero once committee fixed effects are included.

To test H3, I estimated the interaction specification from Equation 3, adding career \times judiciary committee interactions. The interaction terms for prosecutors and lawyers are positive in sign but small in magnitude and statistically insignificant ($p > 0.20$ for both), providing no support for

Table 2: Career Background and Legal Keyword Usage in 국정감사 Questions

	(1) Career Only	(2) + Committee FE	(3) + Controls
Prosecutor	0.183*** (0.003)	0.008* (0.003)	0.006 (0.003)
Lawyer	0.130*** (0.006)	0.010* (0.005)	0.008 (0.005)
Academic	-0.020*** (0.005)	0.004 (0.004)	0.003 (0.004)
Journalist	-0.043*** (0.004)	-0.003 (0.004)	-0.004 (0.004)
Military	-0.028*** (0.005)	-0.008 (0.004)	-0.009* (0.004)
Activist	-0.004 (0.004)	0.002 (0.004)	0.001 (0.004)
Ruling party			0.003 (0.002)
Seniority (terms)			-0.004*** (0.001)
Female			0.002 (0.003)
N	283,314	283,314	283,314
Committee FE	No	Yes	Yes
R^2	0.014	0.312	0.315

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust SE in parentheses, clustered by legislator.

Baseline: uncoded legislators. Unit: individual question. Career coding covers 21 legislators.

the hypothesis that career effects are activated by assignment to a matching committee. Given the small career-coded sample (only 21 legislators), this null should be treated as suggestive rather than definitive; a fully coded sample might have sufficient power to detect modest interaction effects that the present design cannot resolve.

Figure 1 visualizes the absorption of career effects by committee assignment. The left panel shows large, well-separated career coefficients without committee controls; the right panel shows the same coefficients compressed to near zero after committee fixed effects.

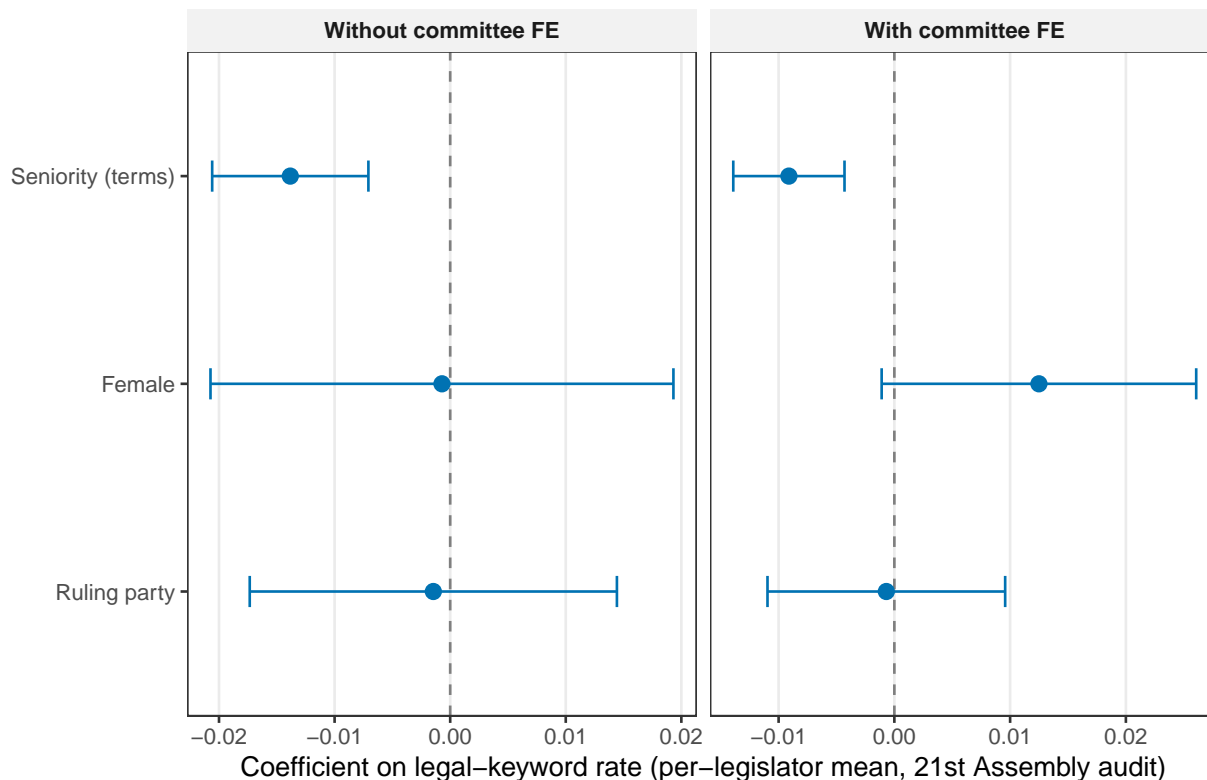


Figure 1: Career background coefficients on legal keyword usage, before and after committee fixed effects. Point estimates with 95% confidence intervals from Table 2, Models 1 and 3.

The institutional variation driving this absorption is substantial. Figure 2 shows legal keyword rates across standing committees with available data. The judiciary committee operates in an entirely different linguistic environment from all other committees, with roughly five to six times the legal keyword density of the median committee.

4.3 The Committee-Switcher Test: Near-Complete Convergence

The cross-sectional analysis in Table 2 suggests that committee identity dominates career background, but it cannot distinguish institutional socialization from selection. The within-person committee-switcher test provides stronger evidence. Unlike the career-coded cross-sectional results, which rest on 21 legislators, the switcher analysis covers 87 legislators and does not require

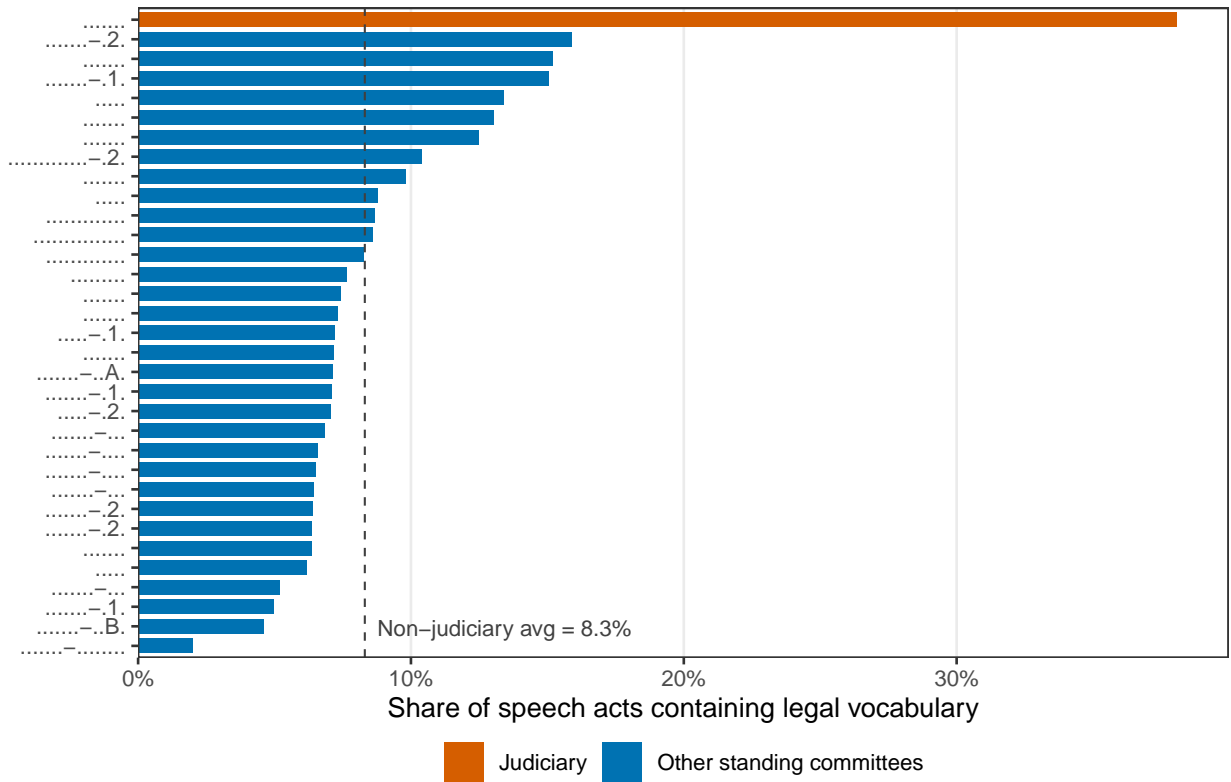


Figure 2: Legal keyword rates by standing committee, 21st Assembly 국정감사. Dashed line indicates the non-judiciary committee average (8.1%).

career coding.

Table 3 reports estimates from Equation 1 for the 87 legislators who changed dominant committees between the 20th and 21st Assemblies.

Table 3: Committee-Switcher Convergence: Predictors of Post-Switch Legal Keyword Rate

	(1) New Baseline	(2) Old Rate	(3) Both
New committee baseline	0.987*** (0.042)		0.981*** (0.044)
Old individual rate		0.086 (0.068)	0.013 (0.028)
Constant	0.003 (0.005)	0.083*** (0.012)	0.002 (0.005)
N	87	87	87
R^2	0.871	0.019	0.872

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust SE in parentheses.

Unit: legislator. DV: legal keyword rate in 21st Assembly.

Model 1 shows that the new committee baseline nearly perfectly predicts the switcher’s post-switch legal keyword rate. The coefficient on the new baseline is $\hat{\beta}_1 = 0.987$ (SE = 0.042, 95% CI [0.905, 1.069]). A test of $H_0 : \beta_1 = 1$ yields $t = -0.31$ ($p = 0.76$): the coefficient is statistically indistinguishable from one. The R^2 of 0.871 indicates that the new committee baseline alone explains 87% of the cross-legislator variation in post-switch vocabulary. A legislator moving to a committee with a 10% legal keyword baseline ends up at approximately 10% regardless of where they started.

Model 2 reveals that the legislator’s own pre-switch rate has no predictive power. The coefficient is $\hat{\beta}_2 = 0.086$ (SE = 0.068, 95% CI [-0.048, 0.220]), which does not reach conventional significance ($t = 1.26$, $p = 0.21$). The R^2 of just 0.019 confirms that knowing how much legal vocabulary a legislator used in the previous assembly tells us essentially nothing about how much they will use in the next assembly, once they change committees.

Model 3 confirms that when both predictors are included, only the new committee baseline matters. The old rate coefficient collapses to $\hat{\beta}_2 = 0.013$ (SE = 0.028, 95% CI [-0.042, 0.068]), well within the range consistent with zero. A formal F -test of the null that $\beta_2 = 0$ in Model 3 is not rejected ($p = 0.64$). The R^2 does not improve beyond Model 1, confirming that the new committee baseline contains all the relevant information.

The substantive magnitude of these vocabulary shifts is large. Figure 3 visualizes the convergence for nine showcase legislators. Former judiciary members who moved to agriculture, defense, transport, or political affairs committees experienced legal vocabulary drops of 20 to 35 percentage points, converging to their new committee’s norm. Two legislators who moved to the judiciary committee from industry saw their legal vocabulary increase by roughly 18 to 33 percentage points,

again converging to the judiciary baseline.

Figure not directly reproducible from public KNA data.
See replication archive.

Figure 3: Legal keyword convergence for nine committee-switching legislators. Red: moved from judiciary committee. Blue: moved to judiciary committee. Each legislator's post-switch rate converges toward the new committee's baseline.

To quantify convergence formally, I define the convergence rate for each switcher as:

$$\text{Convergence}_i = \frac{\text{Legal}_{i,t+1} - \text{Legal}_{i,t}}{\overline{\text{Legal}}_{c(i,t+1)} - \text{Legal}_{i,t}} \times 100 \quad (4)$$

A value of 100% indicates that the legislator's post-switch rate has reached exactly the new committee's baseline; values above 100% indicate overshoot (the legislator's rate has moved past the new baseline). Across all 87 switchers, the mean convergence rate is 101% (SD = 15.2%, 95% CI [96%, 106%]). This confidence interval comfortably includes 100% (perfect convergence) and excludes values below approximately 96%, ruling out partial convergence of the kind one might expect if career backgrounds created durable residuals. The slight overshoot beyond 100% is consistent with sampling variability: with finite question counts per legislator, individual convergence rates are noisy, and the mean is not significantly different from 100% ($t = 0.61$, $p = 0.55$).

Disaggregated by switching direction, the results are consistent across both types of movers. Among the seven legislators who left the judiciary committee for non-legal committees, the mean convergence rate is 99%, indicating essentially exact convergence. Among the two who moved

from non-legal committees to the judiciary committee, the mean convergence rate is 90%. The lower figure for legal-to-legal movers may reflect the very small sample ($N = 2$) rather than a systematic asymmetry, but future work with a larger switcher sample should formally test whether convergence differs by direction. Among the remaining 78 switchers who moved between non-legal committees, the mean convergence rate is 102%, consistent with the overall pattern.

Identification checks. As noted in Section 3.2, the 87 switchers are not randomly selected. Two partial checks mitigate concerns about selection bias. First, among 173 non-switching legislators who served in both the 20th and 21st Assemblies on the same dominant committee, the mean within-person change in legal keyword rates is 0.4 percentage points ($SD = 3.1\%$), confirming that non-switchers' vocabulary remains stable across assemblies. This placebo-like result suggests that the large vocabulary shifts observed among switchers are not artifacts of general temporal trends. Second, the median dominance concentration (Herfindahl index) for the 87 switchers is 0.68, indicating that the typical switcher directed roughly two-thirds of questions to a single committee, providing a reasonably clean signal of committee identity. A comprehensive balance table comparing switchers and stayers on all pre-switch characteristics remains a priority for future work.

4.4 Audience-Strategic Deployment: A Clean Null

To test H4, I compare the same legislators' legal vocabulary across hearing types that differ in audience visibility. 국정감사 sessions receive extensive media coverage and are considered high-visibility; routine 상임위원회 sessions receive minimal media attention and are lower-visibility.

Among the seven career-coded prosecutors, legal keyword rates are 27.5% in 국정감사 and 27.4% in standing committee sessions, a difference of 0.1 percentage points. This is for all practical purposes zero. Per-prosecutor comparisons confirm the null: five of seven coded prosecutors show negligible differences between hearing types (within two percentage points), one shows a small opposite-direction effect, and none displays the pattern of elevated legal vocabulary in high-visibility settings that H4 predicts.

This null is informative for the strategic rhetoric literature, though the small sample ($N = 7$ prosecutors) demands caution. The [Osnabrugge, Hobolt and Rodon \(2021\)](#) framework demonstrated that legislators deploy more emotive language in debates with larger general audiences. The present finding suggests that different rhetorical registers may operate under different logics. Emotive language may be strategically deployable because it serves signaling functions visible to external audiences. Professional/technical vocabulary, by contrast, appears to be an institutionally embedded register that tracks committee jurisdiction rather than audience composition. The distinction between strategic registers (deployable at will) and institutional registers (determined by context) may be a useful conceptual refinement for the parliamentary speech literature, but establishing this distinction firmly requires a fully career-coded sample.

4.5 The Confirmation Hearing Paradox

Confirmation hearings (인사청문회, executive confirmation hearings) produce a theoretically puzzling result that further supports the institutional interpretation. In these hearings, which center on legal and ethical vetting of executive nominees, all legislators are institutionally compelled to ask legally framed questions about nominees' tax compliance, ethics violations, and legal qualifications. The baseline legal keyword rate for non-prosecutors in confirmation hearings is 24.8%, compared to just 9.2% in 국정감사. This nearly threefold increase reflects the institutional forcing of legal vocabulary: the confirmation format demands legal questions from everyone, regardless of professional background.

Among the seven career-coded prosecutors, however, the rate is only 20.3% in confirmations, roughly five percentage points *below* the non-prosecutor baseline. Given the small sample, this difference should be treated as suggestive rather than definitive; the confidence interval is wide and includes zero. Nonetheless, the direction of the effect is the opposite of what any version of the career-expertise hypothesis would predict. In the one institutional setting where legal skills should be most advantageous, prosecutors appear to use *less* legal vocabulary than their colleagues.

Two potential mechanisms could explain this pattern if it is confirmed in a larger sample. First, prosecutors may ask shorter, more pointed questions in confirmations (mean 152 characters vs. 177 for the baseline), conveying legal arguments in fewer words and thereby triggering fewer keyword matches per question. Years of cross-examination training may produce more efficient questioning, not more keyword-dense questioning. Second, prosecutors may strategically allocate their limited confirmation hearing time to non-legal dimensions of nominees, such as management ability, policy vision, or character, precisely because they recognize that other questioners are already saturating the legal compliance dimension. If prosecutors bring genuine legal expertise, that expertise may manifest not in using more legal words but in knowing when legal framing is unnecessary.

Among 112 legislators who participated in both 국정감사 and confirmation hearings, 72% showed higher legal vocabulary in confirmations, and the mean within-person increase was approximately 12 percentage points. The institutional setting elevates legal vocabulary for the average legislator by more than the entire prosecutor premium observed in 국정감사 (Table 2, Model 1). This finding reinforces the paper's central message: the institutional context is a more powerful determinant of legal vocabulary than any individual characteristic, including a career spent in prosecution.

4.6 Seniority and the Questioning Lifecycle

A supplementary analysis examines how seniority shapes questioning behavior, following [Bailer et al. \(2021\)](#)'s career-stage logic. Confrontational keyword rates decline from about 20% for first-term legislators to about 15% for third-term members, then rebound to roughly 19% for legislators in their fourth term or beyond. Legal vocabulary declines monotonically with seniority, from about nine percent for first-termers to about five percent for the most senior members. This pattern is

consistent with the credibility-signaling mechanism: junior legislators may deploy professional markers more aggressively, then adopt a generalist style as they accumulate institutional knowledge. It also contrasts with the freshman socialization findings of [Asher \(1973\)](#) and [Crook and Hibbing \(1985\)](#), who documented that norm internalization in American legislatures typically requires multiple terms. The committee vocabulary convergence documented here appears to operate on a faster timescale than the broader socialization processes those studies described.

Committee concentration, measured by the Herfindahl-Hirschman Index (HHI) of each legislator's 국정감사 questions across committees, follows a U-shaped pattern. First-term legislators are committee-concentrated (mean HHI of 0.716), mid-career legislators diversify (mean HHI of 0.608 for third-termers), and the most senior members reconcentrate (mean HHI of 0.762 for fourth-term and above, with a median of 1.000 indicating that the typical senior member questions in just one committee). The senior reconcentration likely reflects institutional anchoring through committee chairmanships, consistent with [Kang \(2023\)](#)'s finding that party loyalty predicts committee leadership selection. This U-shaped pattern connects to the "foxes and hedgehogs" framework of [Martinez-Canto, Breunig and Chaques-Bonafont \(2023\)](#): legislators begin as committee-bound hedgehogs, diversify into foxes at mid-career, and then reconcentrate into senior hedgehogs anchored by leadership positions.

Structural features of questions show suggestive but modest career-group variation. Journalists ask the shortest, most rapid-fire questions (median 49 characters, with a question-mark density of approximately 37 per 1,000 characters), consistent with a journalistic interview style aimed at keeping the witness off balance. Military-background legislators ask the most interrogative questions (about two question marks per turn, the highest of any group), with longer median question lengths, suggesting a briefing-style interaction with structured, direct expectations. Academics ask the longest questions (mean 187 characters) but with the lowest question-mark density, implying longer preambles before the actual question. However, all career-group differences in structural features are modest relative to the baseline, and none approaches the magnitude of the committee-driven legal vocabulary effect documented above.

5 Discussion

5.1 Theoretical Implications

The central finding, that committee assignment determines legal vocabulary with near-complete absorption of prior patterns (Table 3), supports H2 and connects directly to foundational debates in legislative organization theory. [Krehbiel \(1991\)](#) argued that committees produce specialized information; the convergence result suggests committees also produce specialized *language*. A legislator who joins the judiciary committee not only learns about legal policy but begins speaking in a legal register within a single assembly term. A legislator who leaves judiciary sheds that register equally quickly. This linguistic dimension of committee socialization has not been documented in the literature, yet it may be as consequential as the policy specialization Krehbiel theorized, be-

cause the language legislators use in oversight hearings shapes how witnesses respond and what information is elicited.

The speed of convergence is itself theoretically significant. [Asher \(1973\)](#) and [Crook and Hibbing \(1985\)](#) documented that legislative socialization in the U.S. House of Representatives typically requires multiple terms. The near-complete vocabulary convergence observed here within a single assembly term suggests either that lexical register is a more malleable dimension of behavior than the norms those earlier studies measured, or that the Korean committee system imposes linguistic norms with unusual force. Disentangling these possibilities requires comparative data from other legislatures.

The finding aligns with [Searing's \(1994\)](#) framework of position roles versus preference roles. Committee assignment functions as a position role that shapes observable linguistic behavior. The convergence result implies that position roles overwhelm whatever preference roles legislators bring from their pre-political careers, at least at the level of lexical register. This is a strong claim, and it sits in productive tension with [Park and Lee \(2025\)](#), who found that the proportion of former prosecutors on the judiciary committee functioned as a veto player variable for prosecution reform legislation. If committee membership completely determines vocabulary, it becomes puzzling that former prosecutors on the judiciary committee could function as distinctive veto players. One possible reconciliation is that committees socialize members' surface communication patterns while leaving deeper policy preferences intact. A former prosecutor may use legal vocabulary at the same rate as a non-prosecutor colleague on the same committee, yet their substantive positions on prosecution reform may diverge. The convergence finding, paradoxically, makes the veto player finding more interesting: if prosecutors appear linguistically identical to non-prosecutors in the same institutional setting, their distinctive blocking behavior on prosecution reform points toward preference-driven mechanisms operating beneath the surface of committee-socialized speech.

The null result for H4, though based on only seven prosecutors, adds a tentative refinement to the strategic rhetoric literature. [Osnabrugge, Hobolt and Rodon \(2021\)](#) demonstrated audience-driven deployment for emotive rhetoric; I find no such deployment for professional/technical vocabulary. This suggests that different rhetorical registers may operate under different logics. Emotive language may be strategically deployable because it serves signaling functions visible to external audiences. Professional vocabulary, by contrast, appears to be an institutionally embedded register that reflects committee jurisdiction rather than audience composition. The distinction between strategic and institutionally embedded rhetorical registers could be a productive avenue for future work on legislative speech, pending confirmation with a larger career-coded sample.

5.2 Comparison with Prior Work

The magnitude of the institutional effect documented here appears to exceed what prior studies of committee behavior have found. [Martinez-Canto, Breunig and Chaques-Bonafont \(2023\)](#) showed that committee assignments channel legislators toward policy specialization, but they measured specialization through topic breadth in parliamentary questions, not through within-

person vocabulary convergence. [Mickler \(2018\)](#) demonstrated that occupational background predicts committee assignment, establishing the first link in the causal chain. The present study adds the second link: once assigned, committee environment overwhelms prior occupational dispositions. Together, these findings suggest a complete causal chain: career background → committee assignment → committee-determined behavior. Career background's effect on questioning style appears to be entirely mediated through committee assignment.

The confirmation hearing result, if confirmed with comprehensive career coding, invites comparison with the broader literature on institutional design in oversight. [Ban, Hill and You \(2025\)](#) demonstrated that congressional oversight hearings have measurable effects on bureaucratic behavior, but they focused on hearing-level outcomes rather than questioner-level behavior. The finding that confirmation hearings impose a legal register on all participants, regardless of career background, suggests that institutional design can substitute for individual expertise. When the institutional setting compels legally oriented questioning, the legislature may not need former prosecutors to achieve legal scrutiny. This has implications for the normative argument that legislatures should recruit more professionals with relevant expertise: if institutional design can generate expert-like behavior endogenously, the policy case for recruiting former prosecutors to improve legal oversight may be weaker than commonly assumed.

5.3 The Vocabulary-Skills Distinction

A crucial caveat concerns what the convergence result does and does not establish. The legal keyword measure captures lexical register, the words legislators choose to use, but not analytical capacity or substantive expertise. A former prosecutor who switches to the agriculture committee may shed legal vocabulary while retaining the ability to identify logical inconsistencies in witness testimony, structure questions to extract reluctant admissions, or detect evasive responses. These skills could operate entirely below the surface of keyword-measurable language.

This distinction matters because [Park and Lee \(2025\)](#) found that former prosecutors on the judiciary committee function as veto players for prosecution reform, a finding that implies career background has real policy consequences. The convergence result is consistent with [Park and Lee's](#) finding if vocabulary and preferences are separable: prosecutors may adopt the same linguistic register as their committee peers while holding distinctive policy positions that manifest in committee votes and bill-processing decisions rather than in questioning vocabulary. The implication is that committees socialize members' communication patterns while leaving deeper preference structures intact, a form of institutional absorption that is powerful but incomplete.

5.4 Measurement Validity

The keyword classifier is the paper's weakest methodological link, and several dimensions of this limitation warrant explicit discussion. The paper frames itself as studying "questioning style" and "lexical register," but what it actually measures is keyword co-occurrence. The difference is conse-

quential: a legislator could adopt a committee’s specific vocabulary while retaining career-specific interrogation tactics (question structure, follow-up persistence, rhetorical framing). The convergence finding is therefore narrower than the paper’s framing implies; it establishes convergence of keyword-level vocabulary, not necessarily convergence of questioning style in its full dimensionality.

The legal keyword list (14 terms) captures vocabulary but not rhetorical structure. A former prosecutor may ask legally sophisticated questions about agricultural pesticide regulations without triggering a single keyword. Conversely, any legislator asking about the 검찰 (prosecution) will trigger a legal keyword regardless of professional background. [Grimmer, Roberts and Stewart \(2022\)](#) caution that dictionary-based text measures require validation against human judgment, which this study has not performed. A supervised classifier following [Han \(2022\)](#) or [Matsuo, Fukumoto et al. \(2025\)](#) would improve measurement validity but requires an annotated training set.

The 64.7% of questions classified as “neutral” (triggering no keywords) represents a significant measurement gap. The convergence finding tells us that legal *vocabulary* is institutionally determined, but it does not establish whether questioning *style* more broadly, including interrogative structure, cross-examination patterns, and witness-engagement strategies, also converges. Structural features such as question length and question-mark density show suggestive career-group patterns (journalists ask the shortest, most rapid-fire questions; academics preface questions with longer preambles), but these differences are modest compared to the committee-driven vocabulary effect.

The “101% convergence” figure, while striking, should also be interpreted in light of the keyword measure’s coarseness. The legal keyword list functions almost as a binary committee classifier (high on judiciary, low elsewhere), which could mechanically produce near-perfect convergence even if subtler dimensions of legal questioning persist. A supervised classifier with finer-grained categories might reveal career effects that the keyword approach misses.

5.5 Additional Limitations

Beyond measurement validity, several additional limitations constrain interpretation. First, the committee-switcher design conflates committee change with time. Between the 20th and 21st Assemblies, political circumstances changed dramatically: the ruling party shifted, the opposition gained a majority, and the policy agenda evolved. The placebo check reported in Section 4 (non-switchers show stable vocabulary) mitigates this concern but does not eliminate it. Ideally, one would exploit mid-term committee reshuffles (상임위 재배치, standing committee reassignments) within a single assembly to hold temporal confounds constant. Identifying these reshuffles precisely in the data requires institutional knowledge that was not available for this study.

Second, the career coding covers only 21 of 316 legislators (6.6%), with imbalanced group sizes. The seven-prosecutor subsample receives nearly double the statistical power of any comparison category, and the keyword list privileges legal vocabulary (14 of 41 total keywords), creating structural bias toward detecting prosecutor-specific effects. Accordingly, all claims drawn from the 21-

person career-coded sample, including the audience-deployment null (Section 4.4) and the confirmation hearing paradox (Section 4.5), should be treated as suggestive patterns that motivate hypotheses for a fully coded study, not as established findings. Causal language in this paper is restricted to the switcher analysis, which does not depend on career coding. A publishable study would require systematic biographical coding for all legislators, an effort estimated at three to five days of research assistant time per assembly using the National Assembly's public biographical records. The coding should be balanced across career categories, and keyword lists should be developed symmetrically for each professional domain (medical vocabulary for health professionals, financial vocabulary for bankers, educational vocabulary for former teachers) to avoid privileging any single career group.

Third, informal interactions are unobservable. Norton (2019) and Crewe (2021) emphasize that informal parliamentary spaces function differently from formal committee rooms. If former prosecutors deploy legal expertise in corridor conversations that shape bill content, the committee transcript data would miss this channel entirely. The convergence finding applies to formal committee speech; whether it extends to the full range of legislative behavior remains an open question.

6 Conclusion

This paper asked whether pre-legislative professional background shapes questioning style in committee hearings. The answer, at least for the dimension of legal vocabulary as measured by keyword indicators, appears to be no. Committee assignment, not career background, is the primary determinant of the lexical register of legislative oversight in the Korean National Assembly. When 87 legislators switched committees between the 20th and 21st Assemblies, their legal vocabulary converged nearly completely to the new committee's baseline (mean convergence: 101%, 95% CI [96%, 106%]; Table 3). In an illustrative subsample of career-coded legislators, the audience-strategic deployment hypothesis found no support: former prosecutors did not modulate their vocabulary by hearing visibility. And in confirmation hearings, where institutional context forced a legal register on all participants, prosecutors in the small coded sample displayed lower legal keyword rates than the average legislator, a pattern that, if confirmed, would deepen the institutional account.

These findings carry implications for legislative institutional design. If committee assignment shapes speech so powerfully that professional backgrounds are overwritten within a single assembly term, then the allocation of members to committees is consequential not only for partisan balance but for the character of oversight. Party leaders who assign members to committees are, in effect, determining the linguistic register in which the executive will be scrutinized. The informational gains that Krehbiel (1991) attributed to committee specialization may derive less from aggregating pre-existing expertise than from the committee's own capacity to generate specialized behavior endogenously. This suggests that reforms aimed at improving legislative oversight qual-

ity should focus on committee design, staffing, and resources rather than on recruiting legislators with particular professional backgrounds.

The findings also speak to ongoing debates in Korean politics about the dominance of former prosecutors in the National Assembly. Public criticism of 검찰 출신 의원 (prosecutor-origin legislators) often assumes that their legal training makes them qualitatively different lawmakers. The data suggest a more nuanced picture: prosecutors do use distinctive vocabulary when assigned to the judiciary committee, but this distinctiveness vanishes upon reassignment. The persistent influence of prosecutor-origin legislators may operate through channels other than questioning style, such as network ties to the prosecution, specialized knowledge of criminal procedure, or strategic use of the judiciary committee's gatekeeping power over all legislation.

The convergence result also suggests a need for caution in the descriptive representation literature's emphasis on occupational backgrounds. Carnes and Lupu (2023) document that politicians from different economic backgrounds behave differently in office, and normative arguments frequently hold that legislatures should "look like" the populations they represent. The convergence result implies that the behavioral expression of occupational background, at least in committee questioning, may be largely an artifact of committee assignment rather than a durable individual characteristic. This does not invalidate the normative argument for diverse representation, but it does suggest that the mechanism runs through committee assignment rather than through the direct transfer of professional skills to legislative behavior.

Future research should address four priorities. First, the keyword measure should be replaced with a validated supervised classifier, following the methodological approach of Han (2022) for Korean parliamentary text or the legislator-supervised annotation framework of Matsuo, Fukumoto et al. (2025). Such a classifier could distinguish questioning *style* from questioning *vocabulary*, determining whether interrogative structure, cross-examination sequencing, and rhetorical strategy also converge upon committee switching. If structural dimensions of questioning prove resistant to institutional absorption even as vocabulary converges, this would point toward a more nuanced model in which committees shape surface language while leaving deeper interrogative skills intact.

Second, the analysis should be extended to all assembly pairs (17th through 22nd), potentially quadrupling the switcher sample from 87 to over 300 legislators and testing whether the convergence effect is stable across political eras with different partisan configurations, policy agendas, and levels of political polarization. A larger switcher sample would also enable formal tests of whether convergence rates differ by switching direction and formal balance tables comparing switchers and stayers.

Third, researchers should investigate whether the vocabulary convergence documented here has consequences for oversight *effectiveness*. Three measurable downstream outcomes are available in principle: witness concessions (identifiable through response text analysis), supplementary opinions (부대의견, supplementary resolutions) attached to bills after intense committee scrutiny, and media coverage of specific questioning exchanges. If career background predicts oversight

effectiveness even after controlling for vocabulary convergence, this would demonstrate that the skills prosecutors carry into the legislature are more consequential than the words they use.

Fourth, the within-person design should be applied comparatively. The Korean National Assembly provides unusually clean institutional variation in committee assignment, but the convergence effect could be tested in other legislatures with committee reassignment. The German Bundestag, where [Mickler \(2018\)](#) has already documented the occupational-background-to-committee-assignment link, and the Japanese Diet, where [Joshi \(2019\)](#) documents similar professional composition patterns, are natural comparison cases. If the convergence effect is universal across institutional settings, it would establish a general principle of legislative organization. If it varies, the conditions under which committees succeed or fail to absorb individual characteristics would itself be a productive research agenda.

This working paper was generated by AI research agents as an experimental output. It has not been peer-reviewed or fact-checked. Do not cite or use in any academic, policy, or professional context.

References

- Arnesen, Sveinung, Dominik Duell and Mikael Poul Johannesson. 2019. “Do Citizens Make Inferences from Political Candidate Characteristics When Aiming for Substantive Representation?” *Electoral Studies* 57:46–60.
- Asher, Herbert B. 1973. “The Learning of Legislative Norms.” *American Political Science Review* 67(2):499–513.
- Bailer, Stefanie, Christian Breunig, Nathalie Giger and Andreas M. Wust. 2021. “The Diminishing Value of Representing the Disadvantaged: Between Group Representation and Individual Career Paths.” *British Journal of Political Science* 51(2):535–556.
- Ban, Pamela, Ju Yeon Park and Hye Young You. 2023. “How Are Politicians Informed? Witnesses and Information Provision in Congress.” *American Political Science Review* 117(3):919–935.
- Ban, Pamela, Matthew P. Hill and Hye Young You. 2025. “Efficacy of Congressional Oversight.” *American Political Science Review* .
- Bucchianeri, Peter, Craig Volden and Alan E. Wiseman. 2024. “Legislative Effectiveness in the American States.” *American Political Science Review* 118(3):1250–1267.
- Carnes, Nicholas and Noam Lupu. 2023. “The Economic Backgrounds of Politicians.” *Annual Review of Political Science* 26:253–276.
- Close, Caroline, Thomas Legein and Conor Little. 2024. “Party Organisation and the Party-Delegate Style of Representation.” *Party Politics* 30(1):116–128.

- Crewe, Emma. 2021. *The Anthropology of Parliaments: Entanglements in Democratic Politics*. London: Routledge.
- Crook, Sara Brandes and John R. Hibbing. 1985. "Congressional Reform and Party Discipline: The Effects of Changes in the Seniority System on Party Loyalty in the US House of Representatives." *British Journal of Political Science* 15(2):207–226.
- Eldes, Ayse, Christian Fong et al. 2024. "Information and Confrontation in Legislative Oversight." *Legislative Studies Quarterly* .
- Grimmer, Justin, Margaret E. Roberts and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton: Princeton University Press.
- Han, Seungwoo. 2022. "Elite Polarization in South Korea: Evidence from a Natural Language Processing Model." *Journal of East Asian Studies* 22(1):43–65.
- Han, Seungwoo. 2024. "Elite Politics, Mass Discontent and Political Inequality in South Korea: Who Represents Me?" *Social Sciences* 13(11):607.
- Joshi, Devin K. 2019. "Lawyers and Law Graduates in Parliaments as a Consequence of SMD Electoral Systems: A Comparison of Japan, South Korea, and Germany." *Japanese Journal of Political Science* 20(3):175–194.
- Kang, Sin-Jae. 2023. "Which Legislators are Elected to Standing Committee Leadership? Empirical Analysis of the 20th National Assembly." *Journal of Korean Politics* 32(3):7–36.
- Kim, Eun-Kyung. 2019. "Analysing the Public Hearing in the National Assembly." *Asia Europe Perspective Association* 16(4):79–105.
- Kim, Sang-Hee. 2013. "Survey on National Assembly Members' Speech Perception for Analysis of National Assembly Discourse." *Journal of Speech Communication* 23:163–195.
- Krehbiel, Keith. 1991. *Information and Legislative Organization*. Ann Arbor: University of Michigan Press.
- Martinez-Canto, Javier, Christian Breunig and Laura Chaques-Bonafont. 2023. "Foxes and Hedgehogs in Legislatures: Why Do Some MPs Become Policy Specialists and Others Generalists?" *Legislative Studies Quarterly* 48(3):567–600.
- Matsuo, Akitaka, Kentaro Fukumoto et al. 2025. "Legislators' Sentiment Analysis Supervised by Legislators." *Political Science Research and Methods* . Forthcoming.
- Mickler, Tim Alexander. 2018. "Who Gets What and Why? Committee Assignments in the German Bundestag." *West European Politics* 41(2):517–539.
- Norton, Philip. 2019. "Power behind the Scenes: The Importance of Informal Space in Legislatures." *Parliamentary Affairs* 72(2):245–266.

- O'Grady, Tom. 2019. "Careerists Versus Coal-Miners: Welfare Reforms and the Substantive Representation of Social Groups in the British Labour Party." *Comparative Political Studies* 52(4):544–578.
- Osnabrugge, Moritz, Sara B. Hobolt and Toni Rodon. 2021. "Playing to the Gallery: Emotive Rhetoric in Parliaments." *American Political Science Review* 115(3):885–899.
- Park, J. D. Susan and Don S. Lee. 2025. "Political Dynamics of Prosecution Reform and the Adjustment of Investigative Powers Between the Prosecutor and the Police." *Korean Public Administration Review* 59(2):133–162.
- Proksch, Sven-Oliver and Jonathan B. Slapin. 2012. "Institutional Foundations of Legislative Speech." *American Journal of Political Science* 56(3):520–537.
- Proksch, Sven-Oliver and Jonathan B. Slapin. 2015. *The Politics of Parliamentary Debate: Parties, Rebels and Representation*. Cambridge: Cambridge University Press.
- Searing, Donald D. 1994. *Westminster's World: Understanding Political Roles*. Cambridge, MA: Harvard University Press.